UNIVERSITY OF LONDON THESIS

Degree PhD Year 2007 Name of Author TESSA KATE ANDERSON

A SPATIAL ANALYSIS OF ROAD COLLISION HOTSPOTS AND THEIR DRIVER AND CASUALTY PROFILES

A thesis submitted to the University of London for the Degree of Ph.D.

Tessa K. Anderson

Department of Geography
University College London

January 2007

ABSTRACT

This thesis explores the spatial nature of road collisions within London, UK. There is continuing debate amongst academics and road safety professionals alike as to the most appropriate method of identifying high density locations of road collisions and also of identifying the appropriate data and variables that represent population risk in road environments. This thesis adopts a three stage approach in order to create a typology of collision hotspots and the persons most likely to be involved in incidents at them. The first method links the postcodes of drivers and casualties to geodemographic types in order to understand the types of people in London that are more likely to be involved in a collision and to identify where they are likely to reside. The second method concentrates on defining collision hotspots using kernel density smoothing. The selected hotspots and associated variables are then clustered in order to create a typology of hotspots, using five groups and fifteen clusters. The third and final stage links these two spatial locations together, by ascertaining the geodemographic types which are more likely to be over represented in each of the five groups and fifteen clusters. This makes it possible to develop a clearer and crisper analysis of the road collision risk of the population in London, to classify the capacities in which people are likely to be involved in a collision and to identify where the collision is likely to take place. The outcome is fifteen distinctive types which highlight risk groups across London, suggest the types of collision that they are likely to encounter and identifies where they might occur. This has the potential to be a very useful tool in assisting in road safety policy and initiatives across London.

ACKNOWLEDGEMENTS

There are lots of people I would like to take the opportunity to thank for helping with my research in many different ways. Thanks firstly to my supervisor, Professor Paul Longley who has provided common sense, knowledge and cracking-of-the-whip, without which I would never have finished. I would like to thank him for his continuing support and assistance in relocating to New Zealand. My second thanks goes to Professor Richard Webber who provided much needed guidance and inspiration of tackling a challenging analysis of road collisions. Richard provided insight into both using geodemographics and the clustering technique used. I am extremely grateful for the time he spent assisting me with my PhD analysis and providing much needed direction.

I would also like to take this opportunity to thank the wider academic community at UCL with assistance at various times, especially Professor Gloria Laycock who took time to read and comment on the chapters: her advice proved very useful and insightful. Professor John Adams in the UCL Geography Department has been very supportive and I would like to thank him for his continued interest and help over the past seven years whilst I have been at UCL. I would also like to thank Spencer Chainey for helping me obtain the data for my thesis through the Metropolitan Police, without which, the thesis would not have been written.

I would like to acknowledge my peers within CASA, who provided so much help, advice and support throughout my four years at CASA. These include Dave Ashby, Aidan Slingsby, Christian Castle, Alex Singleton and Andy Crooks. The time spent in Chandler House with these guys was really rewarding, not least because of the much needed technical assistance. They also made the PhD process a lot more fun and interesting, not least at conferences and teaching. I would like to thank them all and wish them all success, in some cases with the Ph.D. write up in the first instance! Finally, I would like to take this opportunity to thank my parents, who endured eighteen months of me living at home trying to complete my PhD and were always ready to listen to my PhD frustrations and offer their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EQUATIONS

APPENDIX

CHAPTER 1

INTRODUCTION

'throughout the world hundreds of millions of motor vehicles mix with billions of people. A pedestrian crossing a busy street tries to make eye contact with the approaching motorist. Will he slow down? The motorist tries to divine the intentions of the pedestrian. Will he give way? Shall I? Shan't I? Will he? Won't he? As the distance between them closes, signals implicit and explicit pass between then at the speed of light. Risk is perceived as risk acted upon. It changes in the twinkling of an eye as eye lights upon it'.

Adams 1995

The analysis of road traffic collisions is not easy, due to their complexity. This thesis presents an original application of techniques which utilize the aspect of the spatial and temporal dynamic which binds road collisions together. Road safety is of great concern to the UK government highlighted by the most recent and influential report 'Tomorrow's Roads: Safer for Everyone' (Department of Transport 2000) which outlines the current trends and statistics in road collisions. The report offers projections for the future and a best practice policy guideline to reduce these numbers. Every year approximately 3, 500 people are killed on Britain's roads and 40, 000 are seriously injured (Mountain, Fawaz and Jarrett 1996). The American Automobile Association estimates that road traffic collisions claim a life every 13 minutes in the USA and the WHO estimates 1.18 million people were killed in 2002 in a road collision which is 2.1% of the global mortality (WHO 2004).. Road traffic collisions have been argued by the WHO to be the leading injury related cause of death among people aged 15-44. Road traffic collisions have formed part of our everyday lives. Every person is at risk, even if they are not a vehicle driver, they are potentially pedestrians, cyclists and at some point every person is subject to using the road network and therefore put at risk from being involved in a road traffic collision. One of the key factors to understanding the complex actions which build up to form a road collision is risk. Risk plays an important role in determining severity, behavioural preventative measures and different types of reaction.

The importance of this research rests on the notion that the traffic system will be acknowledged as the framework within which collisions occur, however what the research will ultimately challenge is that the causal factors surrounding road collisions occur on scales which can be referred to as the 'global' and 'local' effects. Before this is explained it is important to understand what is meant by traffic system. A traffic system is an integrated structure to achieve simultaneous flows of traffic movement, it usually occurs in large urban areas (this research will be focused on London). It suggests a duality between land use and urban transport and striving to make traffic flows more efficient for the user. Due to changing society and its changing needs it is important for some attempt to be made to develop economic, environmental, land use, population and transport planning policies. This research attempts to fuse together these variables.

The study area which will be used for this research is London, UK. There are several reasons for this choice of location. Firstly, the location of the research at University College London meant that pre existing links with the police could be utilised to obtain road collision data. Secondly, offers a large study area of approximately 7 million individuals which can be used as the base population therefore being able to anticipate more accuracy and the ability to differentiate 'risky' people and areas. Although London is inherently a unique city with its road infrastructure and business and population domination within the UK it shares many similar neighbourhood characteristics with other large urban areas such as Birmingham, Leeds, Manchester and Liverpool. This means that, in so far as this research is limited to London, it is possible that similar methods may be applied to these other urban agglomerations. London, therefore was deemed an appropriate and sound choice for the placing of this research. It has the largest proportion of ethnic minorities, of which there has been little research into the effects of road collisions on their population.

It is difficult to overestimate the complex nature of road traffic collisions. Previous research has established a wide range of causal factors included but not limited to drivers mood, and behaviour, weather conditions, passengers unsafe activities, incomplete or dangerous road structure, speed, alcohol and drug use. Much of the present literature and research focuses on specific causal factors which can be limiting in their understanding of road collisions as multi causal events in time and space. With this in mind, it can be hypothesized that people's risk taking behaviour (RTB) is spatially differentiated based on where they live and the socio economic patterns which prevail in postcode and neighbourhood geography. It can be argued this technique is attempting to apply a static solution to a dynamic problem but one that has received little attention in recent years. In other words, the hypotheses for this thesis will be spatially determined and based on variables such as socio economics, collision density and other spatially relevant variables linked to the road environment and individuals involved in the collision. Road traffic collisions occur somewhere and they occur at some point in time. Therefore we can see a road collision as a dynamic event caused by one of or a mixture of human behaviour, the road environment and the vehicle or mode of transportation. Analysis is often concerned with attempting to predict road collisions; this can be called 'pre-collision'. Pre collision analysis relies solely on information about road collisions that occurred in the past and therefore there is strong element of 'post collision' analysis. This continuous cycle highlights how complex a road collision is. Below is a diagram (Figure 1.1) which endeavours to explain the integrated traffic system and the complex relationships that influence collision occurrence.

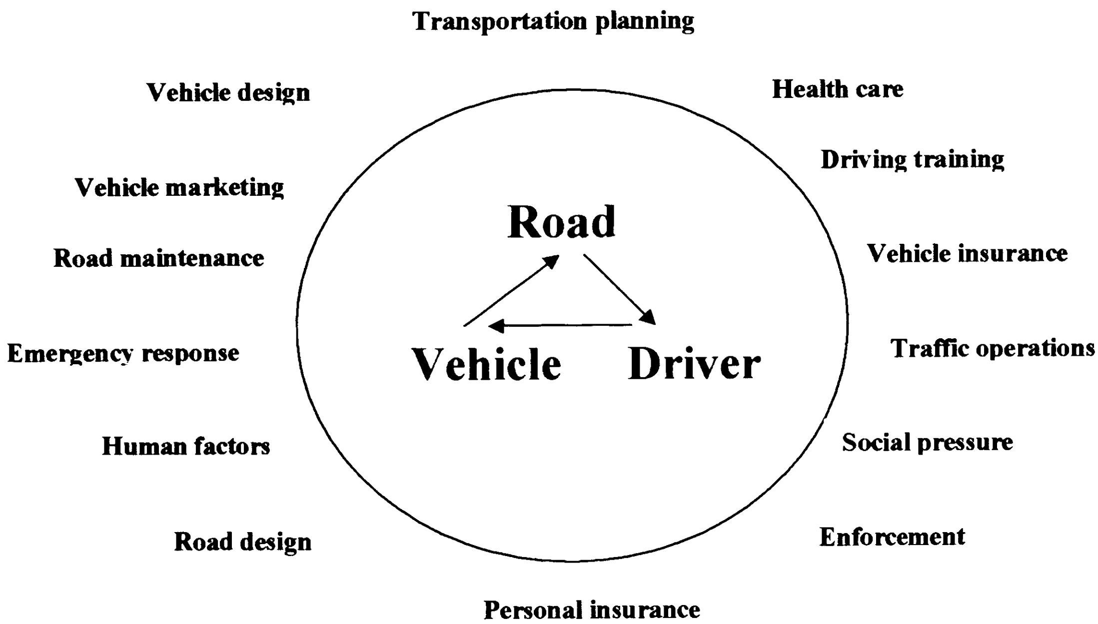


Figure 1.1: *Diagram to show the integrated traffic system (Source: Zein 2002 as cited in Rothe 2002)*

There are two main approaches to road safety and road collision reduction. The first of these is by preventing collision, this means reducing traffic and reducing speed. The second approach to road safety can be determined by the need to reduce the damage that occurs in a collision. Critics have labelled this approach 'safe crashing' and some argue that this approach has been overemphasized by government policies and traffic safety agencies alike (Gladwell 2001). However the backbone of any collision analysis is the datum and its quality, there has been an increasing interest over recent years into the management and collection of road collision data (see Chapter Three).

It has been often said by road safety professionals that data is the cornerstone of all road safety activity and it is ultimately essential for the diagnosis of the road collision problem and monitoring the reduction or management of road collisions. It is important to identify what categories of road users are involved in crashes, what manoeuvres and behaviour patterns lead to collisions and under what conditions collisions occur, in order to define appropriate safety measures. Analysis of road collisions varies considerably and there are no bespoke universal UK guidelines of how road traffic collisions should be analysis, or in other words a best practice guide for prediction and prevention for practitioners and academics alike. For example within London, although all boroughs are managed and funded by the London Accident and Analysis Unit (LAAU), it is the individual boroughs which are responsible for their own area and subsequent analysis and preventative measures.

It is worth noting at this point that there is some division within the literature concerning the definitions of 'accident' and 'incident' and 'crash'. In this thesis the term 'collision' will be used because it is important to acknowledge a vast majority of 'road collisions' are in fact crimes. The word 'incident' does not properly portray the notion that an injury has occurred. The use of 'crash;' is not used because it denotes a too strong meaning. This is also the official wording by the UK police and government alike. A road traffic collision can be defined as 'the product of an unwelcome interaction between two or more moving objects, or a fixed and moving object' (Whitelegg 1986). Road safety and road incident reduction relates to many other fields of activity including education, driver training, publicity campaigns, police enforcement, road traffic policing, the court system, the National Health Service and vehicle engineering.

The field of transportation has come to embrace Geographical Information Systems as a key technology to support its research and operational need. The acronym GIS-T (Geographical Information Systems – Transportation) is often employed to refer to the application and adaptation of GIS to research, planning and management in transportation. GIS-T covers a broad arena of disciplines of which road traffic collision detection is just one theme. Other themes within the discipline of GIS-T include in vehicle navigation systems and Geographical Positioning Systems (GPS). Initially the use of GIS in transportation was only used to query simple collision questions such as depicting the relative incidence of collisions in wet weather or the adequacy street lighting, or to flag high absolute or relative incidences of collisions (see Anderson 2002). Recently there has been increased acknowledgement that there is a requirement to go beyond these simple questions and to extend the analyses. It has been widely claimed by academics and the police that knowing where road collisions occur must lead to better road policing, education, engineering and awareness.

The remainder of the introduction is sub divided into three sections. The first section provides an overview of the present policies coupled with how the history of road safety and collisions has shaped today's thinking by shifting paradigms and differing schools of thought. The second section draws upon the theoretical themes which provide the backbone of this research, which arguably are the nature of risk and how environmental and demographic factors relate to each other in order to explain the patterns and processes of road collisions. The final section presents the aims and objectives for this research based on the importance of road safety knowledge in society and for government policies and the nature of its analysis with reference to geographical theories and hypotheses.

1.1 Definitions

Over the past few years there has been a move away from using the term accident to describe an incident involving damage, injury or death on the road. Instead the police, media and insurance companies have taken to using more descriptive words. The change has been driven by those who suffered damage or relatives of those injured or killed who do not accept the actions or lack of actions on part of the responsible driver. Consequently the word collision or crash are being more widely used. It is seldom found that a single factor was responsible for a collision, but for many there is a desire to isolate a sole cause. More often than not a series of events existed simultaneously to produce the unfortunate result. There are considered to be three key areas which may feature in every collision:

- The road user.
- The road environment.
- The vehicle involved.

A vehicle defect will be contributory in a small number of collisions with a road layout or surface defect being the next greatest contributor but the overwhelming element is the road user and his inability to cope with his or her surroundings. Consequently the following definition of a collision may be more representative:

A rare random multi factor event in which one or more road users failed to cope with their environment

1.2 Overview of policies and government initiatives

1.2.1 A historical perspective

Road traffic collisions have seen a surge in volume which follows closely the rise in automobile ownership up until approximately 1964 which saw a peak of road traffic collisions (Figure 1.2). In recent years the number of collisions has fallen to a plateau while licensed vehicles continue to rise.

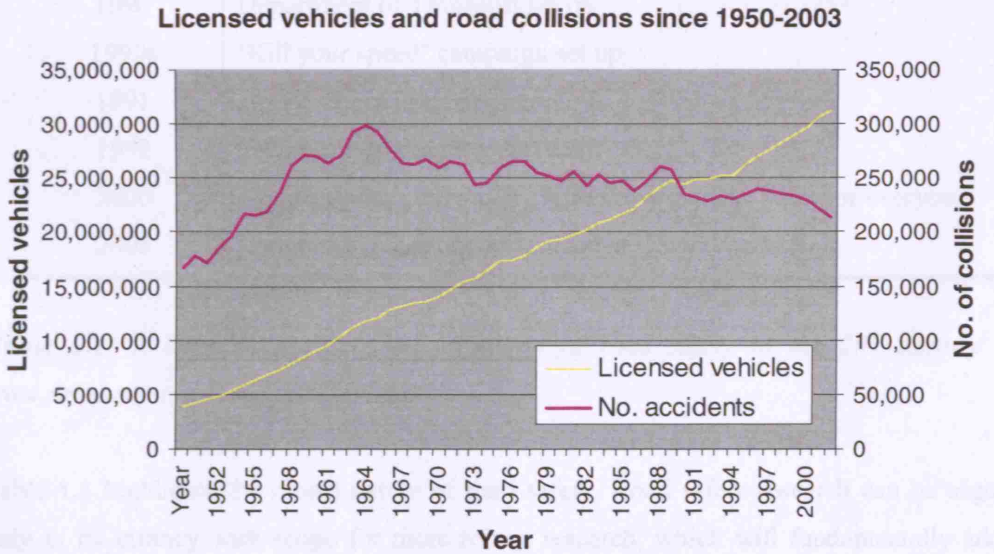


Figure 1.2: *Graph showing the rise in licensed vehicles and road collisions in the UK*
Source: Data from Department of Transport 2003

There have been a number of developments in the road safety domain which shape the current research and policy driven initiatives today, therefore it is useful to reflect on the most important advancements within road safety which can be seen summarised in the table below:

YEAR	*ROAD SAFETY MILESTONE*
1896	First road death recorded, Bridget Driscoll killed by a horse drawn carriage
1899	First fatal road collision involving a motor vehicle
1903	Speed limit increased to 20mph
1919	Ministry of Transport set up
1930	Minimum driving age introduced
1930	Road Traffic Act set different speed limits for different vehicles
1934	Compulsory driving test introduced
1941	Royal Society for the Prevention of Accidents (ROSPA) set up
1957	First motorway opened
1965	50mph limit on certain roads in order to reduce collisions
1975	First roundabout in Croydon
1978	First 'drink drive' campaign
1983	Seatbelts compulsory
1990	Department of Transport set up
1990s	'Kill your speed' campaign set up
1991	20mph zones in urban areas
1992	Speed cameras made permanent
2000	Ten year plan outlined in 'Tomorrow's roads: safer for everyone'
2003	Congestion charging introduced in Central London

Table 1.1: *A brief summary of the evolution of road safety in the 20^{th} century (Source: www.driveandstayalive.com 02.09.05)*

Table 1.1 highlights the recent nature of road safety. Road safety research can be argued to be only in its infancy with scope for more robust research, which will fundamentally address the nature of the geography of road collisions and how they inter-relate to their environment and not just the road environment. Another, perhaps more helpful way of approaching the evolution of road safety is to segregate the various trends of approach to road safety and analysis of collisions, the table below illustrates the shift in paradigms over the past century.

Evolution of road safety paradigms				
ASPECTS	**PARADIGM I**	**PARADIGM II**	**PARADIGM III**	**PARADIGM IV**
Decennia of dominating position	1900 - 1925/35	1925/35 - 1965/70	1965/70 - 1980/85	1980/85 - present
Description	Control of motorised carriage	Mastering traffic situations	Managing traffic system	Managing transport system
Main disciplines involved	Law enforcement	Car and road engineering, psychology	Traffic engineering, advanced statistics	Advanced technology, systems analysis, sociology, communications
Terms used about unwanted events	Collision	Accident	Crash, casualty	Suffering, costs
Premise concerning unsafety	Transitional problem, passing stage of maladjustment	Individual problem, inadequate moral and skills	Defective traffic system	Risk exposure
Data ideals in research	Basic statistics, answers on "What"	Causes of accidents; "Why"	Cost/benefit ratio of means "How"	Multidimensional
Organisational form of safety work	Separate efforts on trial and error basis	Co-ordinated efforts on voluntary basis	Programmed efforts, authorised politically	Decentralisation, local management

Typical countermeasures	Vehicle codes and inspection, school patrols	The three E's doctrine, screening of accident prone drivers	Combined samples of measures for diminishing risks	Networking and pricing
Effects	Gradual increase in traffic risks and health risks	Rapid increase of health risk with decreasing traffic risk	Successive cycles of decrease of health risks and traffic risks	Continuous reduction of serious road accidents

Table 1.2: *Paradigms of road safety (Source: OECD Road Transport Research)*

In order to understand the development of road safety research, it is important to know how the scientific view has changed during the short history of systematic road safety research. It is possible to distinguish four phases of scientific views or paradigms which overlap and interact in a complex way (Table 1.2).

1. Control of the automobile was seen as the problem. There was limited research but more of a description of what was happening. This phase coincided with the rise of the automobile form the beginning of the twentieth century to 1935.
2. Control of traffic situations was seen to be the problem. The countermeasures and the research were centred around the classical three 'E's' approach

- Engineering
- Education
- Enforcement

This is when systematic road safety research was born and when a number of new disciplines came into road safety research. This occurred from 1935 to approximately 1970.

3. Management of the traffic system was seen to be a problem. In this systems approach, mathematical models for description and prediction of traffic accidents were developed. This phase occurred from 1970 to approximately 1985.
4. Management of the transport system as a whole was seen as the problem. The scope is widened from just focusing on the road itself. This is the current trend of road safety thinking.

Road safety research has been studied from top down (aggregate level) and bottom up (individual level) approaches, and this scale is highlighted in the literature. The ultimate purpose of road safety research is to find and implement countermeasure strategies and countermeasure actions that effectively reduce the road safety problems identified. Researchers have however normally focused their interest and efforts on the main road collision variables and to some extent on countermeasure effectiveness. They have rarely extended their interest and efforts to the next stage – how to implement the theoretical and empirical knowledge acquired concerning main road collision variables and effective remedial measures. Figure 1.3 presents the management of road safety and how engineering and behavioural factors are integrated together with the aim to reduce or manage road collisions.

1.2.2 Role of road safety and how it is managed and organized

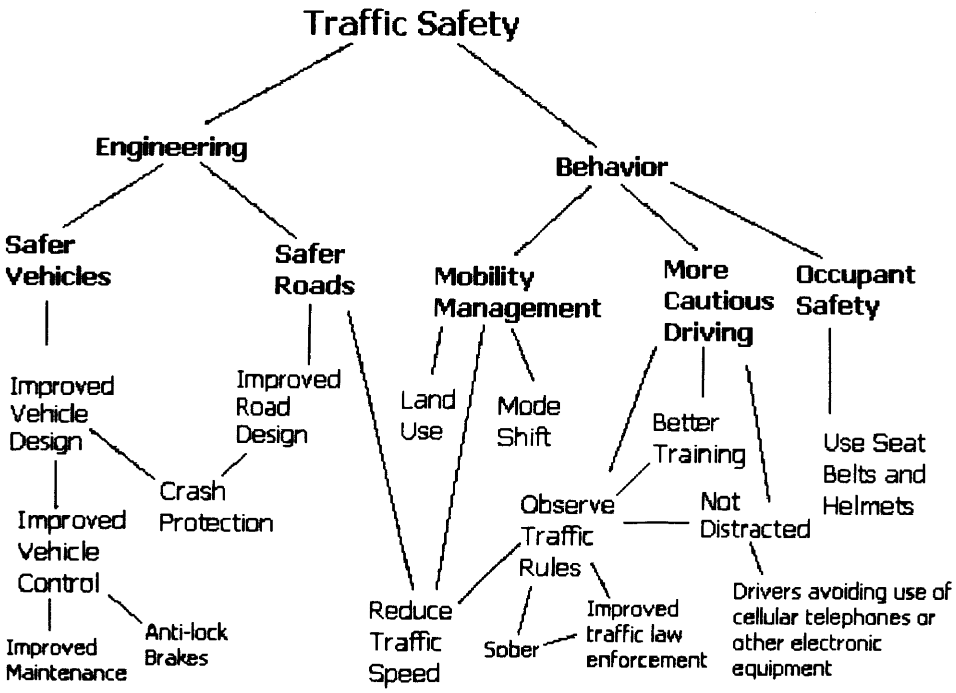


Figure 1.3: *Integrated road safety system, showing the interconnectivity of engineering and behavioural factors (Source: Department of Transport)*

This diagram seeks to act as a general model for traffic safety and how it can be approached and ultimately managed in the countries such as the US and the UK. As road collisions involve roads, motor vehicles and also the human being, the Department for Transport needs to address on a holistic basis, issues covering road engineering, signage, vehicle design, education of road users and enforcement of traffic safety measures. The figure above illustrates the relationships among various traffic safety strategies. The two major categories are *Engineering*, which involves safer vehicles and roadways, and *Behaviour Changes*, which include mobility management (changes in travel mode, route, destination, frequency and speed), more cautious driving, and actions by vehicle occupants such as using seat belts, child restraints and helmets. Wilson and Lipinski (2004) describe many of the engineering strategies for improving traffic safety.

1.2.3 Road traffic collision analysis and preventative governmental policies – a contemporary overview

There is a significant diversity in road policing structures, some centralised, some devolved and no structure was found better or worse (Home Office Research Study 124 1991). One of the major concerns that has come across not only in the government publications but also at 'ground policing' level are the cost-benefits of training police officers to manipulate and understand the data to be able to relate it to every day policing. If the skills are not used on a regular basis then the cost-benefit ratio is greatly reduced. It has proven difficult to persuade and show the policemen and women on the street the importance of data collection and to show what analysis of data can do to enable them to improve policing. In an HMIC road policing report in 1998 a major concern is that of how to measure police performance as far as road policing is concerned. There is required a benchmark standard or measurement with which to outline a level of performance which can be used across the police forces. Figure 1.4 does not mention this aspect but it is important to remember when analysing road collision data and road policing. With respect to this is has been important in recent years to see road policing as an important aspect of policing as a whole and to integrate it into the mainstream policies and objectives, instead of keeping it separate.

The intelligence model (Figure 1.4) shows the needs of road policing, and identifies the need for intelligence led policing and to some extent multi-tasking. The government report concerning road policing identifies the need for the police to re-address the existing balance of resources, which is a point supported by the analysts and academics who believe that there is an increased need for pro-active and problem orientated policing (Goldstein 1991). This is largely because in recent years road policing has been the poor relation of British policing as a whole. The trend is slowly being reversed and this model is trying to increase the recognition that road policing is a critical component of core police work.

The main flaw in the two government reports however is the lack of reference to effective spatial referencing. Road policing is inherently one of many spatial activities within the police and needs to be treated with reference to some theory and effective resource management. In much the same way that ambulances are positioned for optimum route accessibility if they are called out for an emergency. It has been clear that the success of speed cameras has incurred a specific spatial element; however this needs to be more widely applied to the nature of road policing as a whole.

It is easy to outline the flaws within the road policing divisions but in order to establish some guidelines and proposals the problem needs to be tackled from the bottom and this starts with better data collection techniques and more emphasis in police forces on road policing. It should not been seen as a task that encompasses only catching speeding drivers and attending the odd severe accident. There are four key activities identified by the intelligence model (see Figure 1.4) and these include:

1. Hot spot management (proactive policing)
2. Targeting identified offenders (e.g. disqualified offenders)
3. Behaviour management
4. Preventative measures (e.g. speed cameras and traffic management systems)

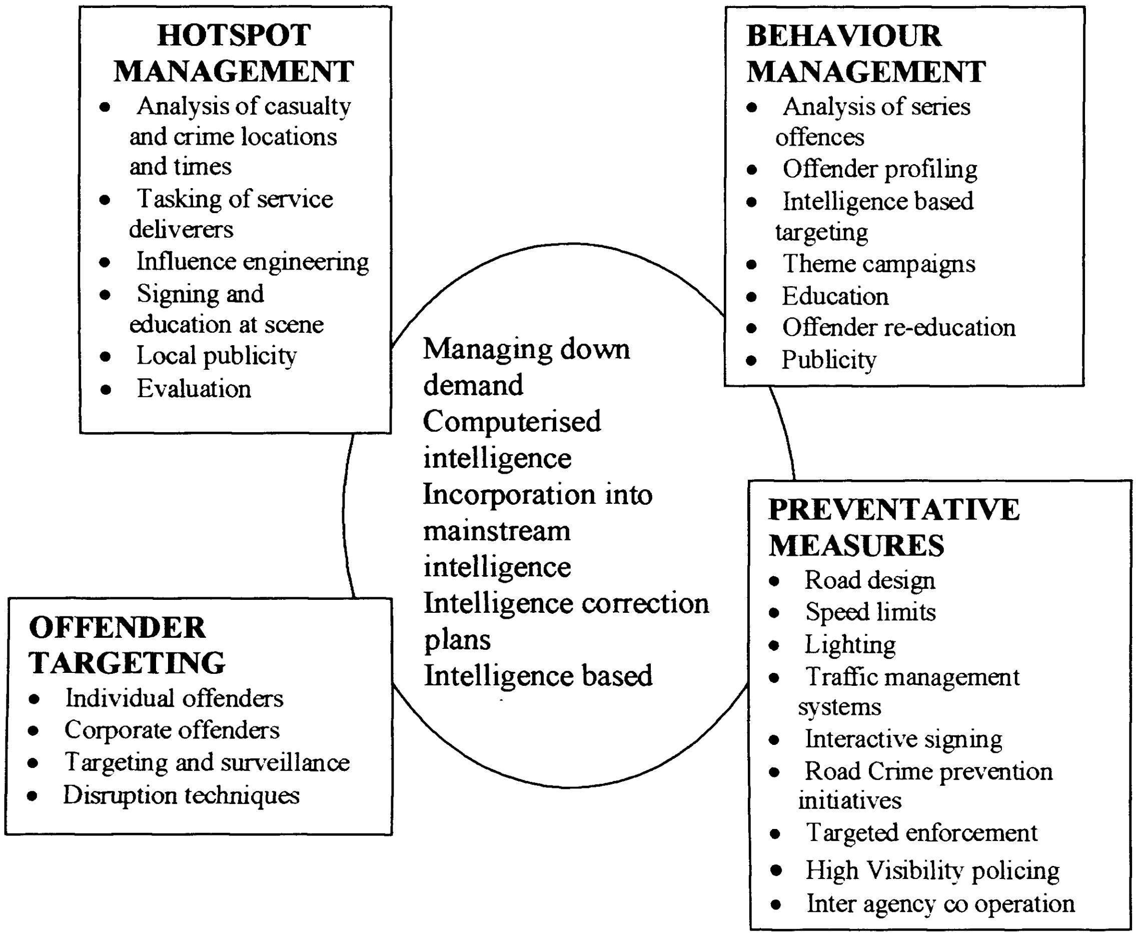


Figure 1.4: *Road traffic intelligence model (Source: Road Policing Report HMIC 1999)*

The significance between all these four points is that intelligence is needed in order to succeed in all four aspects therefore putting more pressure on the police to collect better data and for it to be analysed more efficiently. The 1998 HMIC report referred to previously identifies many police forces managing road-policing intelligence on a 'piecemeal, ad hoc basis which is often personality driven' (HMIC report on road policing 1998). The idea of partnerships also features heavily in the report and states it is important to distinguish between a partnership and a liaison with other agencies. Partnerships involve two or more agencies jointly designing a common strategy, monitored by or on behalf of all parties and often involving the pooling of resources. On the other hand, multi agency liaison involves several bodies, each with its own aims and strategies co-operating with each other because it is beneficial to work together.

As far as targeting certain 'under-achieving' police forces within the road policing sphere it proves difficult because where one force succeeds in reducing the number of fatal accidents another force maybe successful in reducing the number of accidents relating to speeding. The report made many recommendations however the successful uptake of the recommendations depends on the implementation at a local force level.

The Home Office Research Study No. 124 titled 'Traffic Policing in Changing Times' (1991) was conducted over 10 years ago and there have been little or no further studies since which seems to suggest that the role of road policing and specifically the spatial aspect as received very little attention. The main point that has to be acknowledged with regard to road policing is that the police cannot control or reduce every single collision because the cause of the most accidents will be beyond their control. This is with reference to accidents which occur due to road layout, weather or traffic density such as in urban centres. Although it may present itself as a hotspot but there is nothing police would be able to do, apart from work in successful partnership with the local authorities and/or road engineers in order to rectify the problem and enforce speed limits.

In response the research which outlined that policing needed to be either 'policing by objectives' or using 'traffic policing priorities' has seemed to be applied in many of the police forces in England and Wales over the last ten years. Policing by objectives can be defined as 'setting goals and objectives, developing action plans which provide guidelines for control and corrective action' (Lubans and Edgar 1979). As mentioned earlier road traffic police force activities vary from force to force but there should be some universal guidelines for education, training, supervision and communication. The research outlined the amount of time spent on different traffic policing activities. This only supported what academics and police officers already knew which was that the majority of their time was spent completing paperwork. There are so many contentious issues surrounding these activities such as whether the police are stopping the right people, and whether that police action has any significance impact upon driver's subsequent driving behaviour.

As mentioned before many of the solutions to traffic safety lie in strategies such as road engineering which is out of their immediate control (the police). The impact of this kind of police activity on road collisions figures is extremely difficult to measure. In the concluding remarks of the study it outlines the options for traffic policing problems, which consist of:

- Police officers needing to be more aware of the wider economic, social, political and environmental issues concerning management
- Physical preventative measure, i.e. that the roads need to be physically safe
- Technology
- Extra police man power

1.3 Related theories

1.3.1 The nature of risk

Traffic safety professionals can only prevent collisions if they understand what causes them. Traditionally this understanding has been encompassed into the description of a traffic system that consists of three components; the driver, vehicle and the road (or environment). A collision is described as a failure in the interaction between these three components. Police collision data can be aggregated to determine the most common causes of collision. However limitations lie in this process, for example the police are looking for the most direct and immediate cause of the collision, however the cause is often more complex than that, with more overlap between the driver, vehicle and environment. It is a complex dynamic of operators that play a role in creating the circumstances of a collision situation.

In his influential 1965 book, Ralph Nader wrote:

'our society knows a great deal more about building safer machines than it does about getting people to behave safely in an almost infinite variety of driving situations that are overburdening the drivers' perceptual and motor capacities... Vehicle deficiencies are more important to correct that human inadequacies simply because they are easier to analyse and remedy...whether motorists are momentarily careless or intoxicated, or are driving normally when they are struck by another vehicle is entirely irrelevant to the responsibility of automobile makers to build safer cars' page 67.

It is clearly important to understand how people behave on the road and in the traffic environment and why they behave the way they do. By understanding road user behaviour it will provide the government and road safety advisors with the assumptions that underlie the way they design safety programmes and allocate resources. There are several different approaches to understanding the road user. By using the term road user I am encompassing all types of people including pedestrians, cyclists, and motorcyclists. Although the common assumption is that the driver of a vehicle is often the focus of analysis in terms of behaviour, this approach is changing as researchers and officials realize that increasingly pedestrians, cyclists and motorcyclists are the leading cause and casualty in many road collisions. In other words it is not the driver who is at fault, it is however the driver of a vehicle who can inflict more damage and injury on the collision. The important point to understand about risks and the road user is that either we take risks because we don't realize we are taking them or because we are willing to accept them.

Lonero (2000) outlines two different approaches to understanding what I will call the road user. The first one is the 'human factors' or 'ergonomics' approach. This sees the driver or road user as the information processor and performer of set of skills. It fundamentally views a positive side to human nature, and believes that occasionally the demands of a driving task exceed our capabilities. The ergonomics approach looks at what can be altered or managed to vehicles and roadway environments in order to reduce the demands on our limited perceptual and mental capabilities. The second approach is that of the economist and criminologist which sees the human behaviour of the road user as inherently negative. Therefore it is our own actions which determine the difficulty of the task and the risk we experience, in other words our motivation is more important than our capacities and limitations in contributing to the risk. It is important to identify these different views of human nature in relation to risk taking and road collisions, as the research presented in this thesis deals with issues surrounding road user risk, behavioural types and ultimately human nature within a spatial context. By trying to predict what type of reaction or collision someone will have and where they are more likely to live and where the collision is more likely to be a better understanding of road user behaviour can be established.

The road environment is fundamental to the nature of risk taking; it can be understood in its simplest terms as man/woman interacting with his/her environment, or nature and culture. Nature in this case being the human being and culture being the machine and created environment it works in. Researchers often talk about the 'road environment' and there has been a significant amount of research concerning the types of environment which are more conducive to higher risk taking by a number of different road users. Different road environments include built up areas, residential areas, high streets, high speed country/rural roads. By linking the type of road environment and the road user it is possible to form a prediction of where collisions are more likely to occur to certain people. The road environment profoundly affects how one drives, to a certain extent this is controlled by laws but a certain percentage of a driving behaviour is governed by our own personal choices of how we chose to drive. This is turn changes according to many different factors including mood, time of day, road surface, light, and performance of the vehicle.

In recent years it has become apparent that a certain degree of 'risk compensation' is taking place (Adams 1995). The hypothesis of risk compensation implies that persons experiencing a real or perceived change in the riskiness of an activity will alter their consumption of that activity to obtain a preferred combination of risk and reward (Adams 1995). It is perceived by many analysts that road users will start to take more risks to compensate for vehicle and roadway improvements. A pedestrian stepping out onto a road will see traffic lights up ahead to slow the driver down, he or she will overestimate the braking time a driver will need or perhaps the pedestrian will cross at an point in the road which is not seen by the approaching driver. All these factors which are occurring at one time will be subject to risk compensation. This term goes under different aliases; risk compensation, danger adaptation, risk homeostasis and moral hazard.

The main challenge facing road collision analysts is determining how a collision occurred using the data available. I have mentioned that a collision is a complex multi faceted event which ultimately can be deemed a 'chaotic' event. To understand the turn of events that led up to the collision is difficult to quantify. It is easy for statisticians to portray collision statistics and the government can determine whether they are increasing or decreasing but to understand the process of how, why, to whom and where is what this research hopes to build on.

1.4 Importance for using GIS for road safety analysis

This section seeks to outline and interpret the relative importance and benefits of using GIS for road safety and more directly road incident analysis. What follows includes methods which are unique to GIS and how they integrate into the methods of this PhD. GIS has been employed to relate, organize and analyse road traffic incidents worldwide. It is clear however that GIS cannot replace the need for the local analyst to interpret the results and recommend engineering, education and enforcement improvements. This section seeks to outline an approach that underpins the benefits of using GIS for road collision data as opposed to using merely the data in a statistical package and onsite identification of road incident causes.

It therefore raises the question; what additional benefits can GIS provide that we do not already exist without in terms of road safety analysis provision? This question is important for this research as it captures a question asked by many local authorities and police who use software to analyse road incidents that is not conventionally classified as 'GIS'.

One of the most common uses of GIS in road safety is to visually digest a large amount of information quickly for example showing a map of high frequency road incident locations. A study in North Carolina used a 'sliding scale' whereby a segment of a specific length along a road was dynamically moved until that segment met a threshold such as a minimum number of crashes or crashes of a particular severity. In this case the threshold can be varied. The task of studying road incidents in a GIS may be represented as a spatial analysis problem. There can be two key benefits which can be deemed from a visual representation of collision locations:

1. An understanding of any clustering of high collision locations
2. Visual patterns may be used to discern geographic and spatial relationships based on selected variables such as drivers age. However this can be narrowed down to a specific query by looking for example at those collisions that occurred Friday and Saturday evenings between 9pm and 6pm that involve male drivers under the age of 24, therefore certain types of problem can be identified.

This thesis approaches both of these collision types, in Chapter Five, the nature of the collision hotspot is addressed and the second point here addresses the key findings from Chapter Four. Austin *et al* (1997) state that it is other types of inquiry that makes better use of GIS. The first of these is error checking where the features on the database can be compared to the features of the incident coded by the attending police officer (for example differences in speed limit from database to incident report). The second aspect of GIS usage is to identify collision regions or zones as opposed to identifying specific intersections or segments. This allows the analyst to categorize areas by land use and compare how they affect the number and spatial layout of incidents. An example application would be analysing child pedestrian safety on the route to school. It would mean an in-depth analysis could be made of the neighbourhood and an evaluation of routes to school and their relative safety. These points that Austin el al (1997) makes are extremely relevant to explaining the motivation behind this PhD and how different variables and information can be knitted together within a GIS and statistical databases in order to build up as much information as possible regarding the circumstances of the collision, including pre-collision, at scene collision and post collision information.

1.4 Challenges for road collision research

Old approaches to road collision research emphasized the concept of problem-solving in road safety, but it is more correct to recognize that road safety activities don't solve problems. For instance, when a safer road design is implemented, hopefully the number of crashes, or their seriousness, will go down, but they will not disappear. It is more correct to see crashes as an area where the implementation of correct policies, programs and measures will reduce his numbers or consequences, but they will no be ′′solved′′.

This realization is important, because it changes the focus from a problem that will go away if we devote enough resources to it, to a situation requiring on-going management. This management in turn requires the development of scientifically-based techniques, witch will enable us to predict with confidence that safety resources are well-spent and likely to be effective. Therefore with this in mind I have summarized some challenges to road collision research below:

i) Tailoring data management, analysis and especially visualisation of results to the requirements of the user.

The range of agencies requiring information is broad and purposes to which the information applied varies. At a national level, these include government, academics and researchers, organisations working in the field, the private sector and the media. At a regional scale interested agencies include Local Health Authorities (LHAs), for health promotion, planning and preventative work, voluntary sector agencies, inter-agency groupings and local practitioners.

ii) Collating information on the geographic distribution of potential socio-economic, demographic, behavioural and environmental risk factors for comparison with accident distribution.

It has been suggested that there are gaps in the data available concerning causal factors (behavioural and environmental) which lie behind the occurrence of accidents as well as information linking accidents with socio-economic profiles and characteristics. Furthermore, local studies have shown wide differences between accidents occurring in different districts, which can be explained geographically, environmentally and sociologically.

iii) Integration of differing base denominator data

Problems in analysis have been caused by incompatibility of coding systems, use of different populations and denominators and lack of temporal continuity. Standardisation is required to gain a truer picture of trends in casualties (numbers of casualties per unit population) and type of accident (e.g. number of road accidents per unit traffic volume).

iv) Temporal analysis for intervention management

Within accident analysis it is considered essential to review trends and plot changes over time. This is required to examine whether intervention measures are successful and to manage resources for future prevention schemes.

v) Modelling

Estimates of the overall picture can often only be made by extrapolating the findings of local studies.

1.6 Research design: a brief summary

1.6.1 Aim of thesis

The aim of this thesis is to merge the multi casual and multi locational (the 'home' and the 'collision location') dynamic of a road collision to create a taxonomy of road collisions based on variables associated with collision location and home location (based on postcode data) of the driver and casualty. It endeavours to assess and quantify road users in terms of their socio economic characteristics and apply that information to collision hotspots in which the individual was involved. Firstly, the locational analysis for the drivers and casualties will be conducted using postcode data and appending it to geodemographic classifications. Secondly, the collision hotspots will be determined using a novel and innovative technique based on kernel density smoothing in order to determine the most significant variables associated with the hotspot and minimize error. This dualistic information will in turn, provide a classification prediction method for determining 'types' of people in relation to the 'types' of road collision they are more likely to be involved in. Thus ultimately, the aim is to design a risk index which will fuse together the notion of the spatial and the temporal and offer a best practice technique for deriving the index from point of collision to location of driver and victim in order to establish a road safety risk index for people living in London.

1.6.2 Objectives

To achieve the aims of this thesis the objectives are outlined below. The structure of this thesis is such that each objective is addressed sequentially in each separate chapter. Thus each of the following objectives provides an overview of the chapters of the thesis.

Section A – Casualty and driver profiling

1. Using geodemographics to understand the nature of the different and redefined traditional approaches to types of road user risk; using different 'types' of road collision (including information regarding contributory factors)
2. Non spatial analysis, sum by geodemographic classifications, indexing and standardisation of results.
3. Spatial analysis of road user risk profiling, to include buffering around central London, to include maps of different areas of London to highlight communities of certain types with a high propensity of collision risk, and what LEVEL of risk.

Section B – Hotspot profiling

1. Using kernel density smoothing and grid cells to determine high density hotspots using the density of the collisions.
2. Discussion of bandwidth and search radius in order to obtain the optimum results for the hotspot analysis
3. Extracting the hotspot information, with regards to spatial location and information about the hotspot in terms of collision and environmental characteristics.
4. Using a k-means clustering methodology to profile all the hotspots and their characteristics in order to create cluster of hotspots which are more similar in order to compare 'like with like'.
5. These clusters will become hotspot 'types' and be outlined and explained in the text.

Section C – Hotspot typology

1. The hotspot typology will be based on the idea of being able to predict two points of data, firstly by knowing a London postcode being able to predict the most prominent type of collision the person is likely to be involved in and the locational attributes of the location.

2. By turning the data around and using the collision hotspot location data, the results should be able to predict from a collision hotspot type the prominent type of person involved in that type of collision.
3. As mentioned before these predictions will vary across time and space and will create a complex road collision risk prediction strategy for London and its residents.
4. Determining who is more likely to be involved in a certain type of collision.

It is important to note at this stage that these methodological aims and objectives will be discussed in greater depth at the end of chapter two, when a discussion of the hypotheses to be test will take place with specific reference to current literature and research. This section will explain the rationale for choosing this method briefly outlined here and why the approach mentioned here has been taken.

The outline of the thesis is as follows. Chapter Two presents the findings of a literature critique for road safety literature; it particularly discusses the role of geography within spatial road collision analysis and addresses the complexities that occur within this analysis including causal factors and safety policy. Chapter Three comprises of the final section of this introductory section by addressing the nature of scale and uncertainty within road safety analysis and particularly with regards to the scale of collision hotspots and how this is challenged in various studies. The second section concentrates of the three key methodological themes. Chapter Four focuses on profiling the victims (casualties and drivers) of road collisions according to their postcodes using geodemographics in order to determine the types of people more likely to be involved and where they might live in London. Chapter Five in comparison determines the most appropriate method for the road collision hotspot identification using kernel density smoothing and in turn provides the basis for the collision hotspots to use in the clustering process. Chapter Six is the final Chapter in this section and builds on Chapter Five in so far as it takes a proportion of the highest density hotspots and exact collision data for them (428 hotspots) and then uses a k-means clustering algorithm to cluster the hotspots into five groups and fifteen clusters. Chapter Seven in the final section of the thesis brings the two locational datasets together. It takes the road collision clusters and groups and determines the geodemographic types likely to be involved in the clusters and groups. This creates the basis for the typology and enables a prediction, based either on the spatial location of the collision hotspot or where a person lives. Finally Chapter Eight summaries the main findings of the thesis and investigates areas of further study and policy recommendations.

CHAPTER 2

A REVIEW OF ROAD TRAFFIC COLLISION RESEARCH

2.1 Introduction

Road collisions are a persistent consequence of the mobility of today's society. The impact of road traffic collisions in terms of injuries, impairments and fatalities is a major social and public health challenge. Results from road safety research are the most effective means of road collision countermeasures. The overall aims of this chapter are to identify, collate and review published research and other information relating to road collision analysis and suggest how it might be applied to the London wide analysis in this study. Specific aims for this chapter include:

- Highlighting the main contributory factors to urban/London road collisions (infrastructure and road design, socio economics, behavioural and environmental)
- Highlighting particular characteristics of the population which are more likely to be involved in collisions.

Due to the large scope of road safety and collision research overlapping many different themes it has been necessary to narrow the field in order to understand the specific interests needed to apply context to this particular study. This literature review utilizes research from an international arena, predominantly focusing however on urban areas of analysis. The London-wide scale of analysis has been focused on government and borough initiatives and specific studies in this area. This is because there has been limited scope for studies of causal factors for collisions and hotspots within London. This literature review is focused on four key areas:

- The first is to understand the causal factors (risk evaluation) associated with collisions and the spatial dependence of collisions.
- Secondly, infrastructure, the difference between rural and urban road collisions and land use.
- Thirdly, temporal analyses of road collisions which include weekday versus weekend collisions and collisions in darkness or during daylight hours.
- Finally environmental, particularly weather and road surfaces.

A large range of literature exploring causal factors exists, by way of behavioural explanations. This is focused on physiological explanations for collisions as events and will not be reviewed in this chapter. However it is acknowledged as an important body of thought for the explanation and reduction of road collisions and associated hotspots. The first section of the chapter addresses the wider subject of risk associated with road collision involvement. The second section reviews the London wide research initiatives which have been implemented in recent years and how this study will contribute to this important aspect of policy formation through research. The final section evaluates the different methods for road collision analysis with a focus on the geographical aspect of collisions as geographical events in space.

2.2 Context

Road safety involves three major components: the road system, the human factor and the vehicle element. These three elements are inter-linked through geo-referenced traffic collisions (from the police) and provide the basis for road safety analyses and attempts to reduce the number of road traffic collisions and improve road safety. Although numbers of deaths and serious injuries are back to approximately the 1950s levels when there were many fewer vehicles on the road, there are still over 3000 fatalities nationally and this is a considerable squander of human capital. Geography as a discipline in recent years become the over arching theoretical framework to underpin locational and temporal based theories in road safety. The quantitative revolution within geography in the 1970s prompted the focus on the spatial and quantitative significance of events. Road collisions have a geographical place in space and the computational advances in

GIS have enabled people to understand the locational based patterns of road collisions at many different levels.

2.2.1 The geography of road traffic collisions

It was the 1970s which saw the beginning of UK – wide road safety initiatives and analysis. In 1971 the Institute of Road Safety Officers was set up and this year saw the arrival of the Green Cross code. The literature on road safety was beginning to take shape by the late 1960s, but it took many years for the full scope of its research capacity and need for it, to be understood. The early 1980s saw the creation of the Select Committee on Transport (1984) whilst in 1986 it became compulsory for seat belts to be worn and the AA foundation for road safety was launched. This saw the notion of road safety within research increase and become an important driver for road safety initiatives and campaigns.

It has been challenged by Whitelegg (1987) that geographers have not paid enough attention to the geography of road traffic collisions. For many years it was considered by police and local authority road engineers that road engineering, road layout and vehicle manufacturing faults that were the main causes for road collisions. However it has become evident through increasing police awareness that road collisions need to be dealt with in a geographical capacity (both spatial and temporal). In March 2000 the UK Government recorded that every year 3,500 people are killed on Britain's roads and 40, 000 are seriously injured. There was a total recorded average of 300,000 road casualties every year. The report also acknowledges that since the number of vehicles on the road has increased since 1930, the number of collisions has not increased as rapidly (however this is largely to do with vehicle safety technology). Although the number of collisions has not increased in proportion to the number of vehicles it is accepted that a large number of road traffic collisions go unreported due mainly to their not resulting in any injury but still however costing significant amounts of money and time. These types of collisions are addressed in chapter Three and referred to as unreported collisions.

Whitelegg (1987) sought to elucidate road traffic collisions with reference to the scale of analysis and the importance of focusing on neighbourhood and community scale for an answer to the reduction of collisions (see below). Whitelegg's research sought to understand the relationship between road traffic collision analysis and other geographical fields, which include population density, and distribution movement and spatial designs of neighbourhoods. The casualties whether they are pedestrians or motorists will be a function of the land use system, residential patterns, population densities, street geometry, location of workplace, shopping precinct or health centre etc. The themes that come across in this paper are the beginnings of a spatial understanding of road traffic collisions and the understanding of the statistical importance of accurate data. For the purpose of this review this research paper marks the beginnings of understanding road traffic collisions from a spatial perspective in an academic environment (as opposed to government or private consultancy firms).

Increasing spatial scale	**Policy response**
Local/particular	Blackspot eradication/road hump/small scale engineering
Neighbourhood	Alternating residential design
Sector of city	Traffic management and routing
City wide	Public transport system/land use planning

Table 2.1: *Road traffic accident: relationship between spatial scale and policy response (Whitelegg 1987)*

Whitelegg's (1987) work is associated with outlining the importance of a scale based approach when trying to reduce road traffic collisions (Table 2.1) and challenges that human error alone is not responsible for road traffic collisions and the importance of spatial factors have been 'grossly underestimated'.

There is a range of scales at which road traffic collisions can be analysed. For example, local authorities may record collision rates along a specific segment of road to understand the cause in terms of road engineering. The police collect collision data for their authority which can cover many counties for example Thames Valley Police covers more than one county, however this relates more to police geography. It is an important point that due to the variability of scales, it makes the data incompatible and therefore challenging for meaningful comparison.

Whitelegg (1987) outlines his cause for concern as traditional approaches have focused on the human and vehicle element and the immediate road environment. There is little understanding and development of how collisions can be reduced by spatial policies. This perception is consistent with my research which highlights the need to approach the analysis from a neighbourhood wide scale. Traditionally the approach to road collision management has been to focus on eradication of hotspots – areas containing a high number of road collisions either serious or fatal. It is important to understand the road safety dynamics behind the spatial approach to hotspot identification and remedial solutions.

2.2 Risk: an over arching issue of road collisions

The analysis of risk and its role in road collisions has a strong element of human behaviour attached i.e. why drivers take the risk they do. Adams (1999) outlines the strong argument for 'risk compensation' within our society, whereby people perceive themselves as safer or better equipped against danger and are therefore more likely to take risk. This theory has been applied to the use of seat belts and speed. However there are a number of risks which are not apparent to the naked eye whilst interacting the road environment. For example these might include being a certain age; teenagers are a high risk group within our society. In terms of risk and road collisions, risk comes in many different forms (Adams 1999). Although the propensity to take risks and be of increased risk is widely assumed to vary with circumstance and individuals, there is no way of testing this assumption by direct measurement. If a road has many collisions it might fairly be called dangerous but using past collision rates to estimate future risks can potentially be misleading. There are many dangerous roads that may have good collision records because they are seen to be dangerous, and therefore children are not allowed to cross them or elderly people are afraid to drive on them. The good collision record is purchased at the cost of community severance due to road safety intervention at dangerous collision locations in order to make them safer. By the time there is a good safety record, a number of collisions would have already occurred for this intervention to take place. The good collision record gets used as a basis for risk management.

It is clear therefore that people are exposed to different risks in the road environment for many different reasons. For example in a crowded local shopping street on a Saturday morning as cyclists, pedestrians, cars, lorries and buses all compete for the same road space. Not all the dangers confronting the participants are able to be seen. Choices about risk also occur at the community, regional and national level. As a society we decide how much loss we are willing to accept in exchange for how much freedom and mobility. A large number of factors influence what drivers do, ranging from behavioural genetics to visual perception of the economy (Lonero 1998). This section looks at these different risk factors associated with varying scales of risk and perception of it.

2.3.1 The relationship between static and dynamic risk: why it is important?

Some kinds of people are more at risk of being involved in a collision than others (Standish 2003). For example a strong implication of whether someone is more likely to be involved in a collision is their age. In particular, children aged 12-16 are at high risk from being involved in a road collision (www.thinkroadsafety.gov.uk, 2005). The reasons for this increased risk are subject to debate, as it is difficult to underpin the exact causes for collisions, however one in ten teenagers across the UK involved in a collision say they were not paying attention (www.thinkroadsafety.gov.uk, 2005). Road use is highly prone to risk consciousness because other people are perceived as a threat in what have been dubbed our 'risk societies' by Ulrich Beck (1992). The development of risk consciousness is an outcome of profound social change implying that society has problems that cannot be resolved, only managed (Furedi 1997). People tend to think that the risks of driving come from other road users. However transport safety does not exclude our own roles as road users. The key issue surrounding this notion of risk is that when choosing a mode of transport, individuals look towards their own 'perceived risk level' instead of the objective risk level when making their decisions.

The traffic environment is one that is constantly changing. It has been suggested that the greatest factor contributing to collision severity is an underestimation of the level of risk a traffic environment presents. All road safety research places a static risk level or understanding on individuals or areas in what is a dynamic traffic environment. The road traffic environment is constantly being referred to as dynamic and this is because it is constantly changing, varying from second to second. In other words someone's chances of being involved in a road collision regardless of whom they are, where they are from can change within seconds. At an urban city wide scale this static measurement is useful in determining a wide ranging understanding of the risk patterns in a spatial environment.

Often when measuring and trying to manage risk, road safety analysts categorise road collisions and those involved in terms of severity of the collision. This method however according to Adams (1995) does not provide the best allocation of risk measurement. This is partly due to the small numbers of actual fatal collisions that occur, since they are both infrequent and scattered across space and time. Thus in this study, data concerning both fatal and serious collision victims have been merged together in order to create a better indication of the patterns of risk. Another risk inherent in using only fatal or severe collision victim data is the uniqueness of London as an urban road network. Adams (1995) summarises the argument that there is a higher proportion of minor collisions in London compared to the rest of the UK urban road network, and attributes this to the fact that London is so congested and traffic speeds are so slow that there are large numbers of minor collisions but that high speed collisions resulting in more serious injury are more rare. Adams also notes the uniqueness of London's road user risk, as it presents the highest urban UK proportion of cyclist and pedestrian related collisions (1999). This presents a strong rationale for a broad societal risk analysis and evaluation.

2.3.2 Risk factors

Collision risk (or 'accident proneness') has been a long established discipline within the road safety domain. Extensive work by academics such as Wilde (1982), Hauer (1982), Janssen and Tenkink (1987), Adams (1995), and McKenna (1998) have illustrated the propensities for risk within the road environment. A collision has often been likened to a stochastic event, however human participation leads to an elevated element of prediction and anticipation. The pioneering statistical contributions of Greenwood and Woods (1919), Greenwood and Yule (1920), Newold (1926, 1927) and Greenwood (1951), take this approach. The basic thesis was: ''When discarding chance is possible in the statistical data of collisions, then the human factor is the single cause of collisions.'' In a similar vein ''some people have many more collisions than can be expected by chance, so, these people are 'accident-prone' ''.

In the field of traffic research, continued emphasis has been placed on identifying factors that contribute to increased driving risk, with the goal of reducing the frequency and impact of traffic violations, collisions and fatalities. Two major attempts to explain behavioural adaptation to changes in traffic systems were proposed in the 1970s one by Natanen and Summala (1974, 1976) and the other by Wilde (1976) which utilised the concept of risk. The former proposed a threshold model on the assumption that in the dynamic driving situation, drivers actually control safety margins rather than some specific risk measure, and only when the risk or fear threshold is exceeded or expected to be exceeded, does it influence behaviour. They postulated a 'subjective risk' or 'fear monitor' (Summala and Natanen 1988) which alarms and influences driver decisions when safety-margin thresholds are violated. One aspect of this approach is that with repeated confrontations, drivers adapt to situations which at first elicited a 'risk response' and drive most of the time with over learnt habitual patterns based on safety margins, with no concern for risk: hence the label 'zero-risk theory' (Natanen and Summala, 1976; Summala,1988, 1996).

In his 1942 book Why we Have Automobile Accidents, DeSilva noted that 'the degree of hazard to which a driver is subjected, or, expressed differently, the extent of his exposure, is determined by how much, and where, and when he drives' (DeSilva, 1942, p. 11). Thus, it is necessary to split exposure by type of road, visibility, weather conditions, time of day, and day of week and so on. In reducing collisions, we do not always need risk measures (collision rates), but it is practical to search for hotspots in the collision mass - accumulations of collisions which indicate where effort to save lives may be concentrated. The search for hotspots continues to be a major safety activity among road and traffic engineers. Although not always based on proper understanding of causes, and at risk of directing resources on a random basis (e.g. Hauer, 19861) cumulative efforts by road and traffic engineers, together with improved design standards, have gradually made this approach less and less useful.

Along with this 'geographical' hotspot method, however, it is essential to continue searching for hotspots in the collision mass by disaggregating it into smaller units by type, road and traffic conditions, time of week and, where possible, on the characteristics of road users involved. Well-defined hotspots or peaks in the collision mass, which may be widely distributed geographically, are of practical and theoretical significance to collision prevention even without corresponding exposure measures. An example of a major hotspot which is widely distributed geographically is young male drivers' crashes at night on weekends. These hotspots necessarily reflect exposure: both of populations at risk and their behaviour. For example, motivational (life-style) factors on weekend nights especially tend to get young male drivers into problems in contrast to older drivers who appear to avoid impaired conditions (e.g. night driving). The degree of hazard to which a driver is subjected is thus not only determined by how much, and where, and when he or she drives (DeSilva, 1942), but also by how he or she drives - which is related to the level of control over potential dangers. The following section aims to disaggregate these exposure elements into manageable factors which will be discussed in relation to human and environmental elements.

2.3.4 Attempts to delineate high risk road user groups

Four groupings are considered in this section based on the predominant patterns from the research material; young male drivers driving at night, locally born people and immigrants, gender differences and young and old drivers. In terms of identifying these high risk road user groups, much of the literature has concentrated on demographic groupings such as 'new' drivers, however little attention has been focused on the spatial dimension of this issue. For example there are likely to be high risk user groups being present at certain times and at certain locations within an urban area. An example of this limited spatial and temporal perspective was demonstrated in a paper by Pietro (2001) who defined a high risk road user including one of the following:

- New drivers
- Drivers with the infringements that are tracked by the loss of points (e.g. speeding)
- Drivers with criminal convictions
- Drivers with an unusual number of crashes or crash type, in a particular time period or location
- Drivers with certain medical conditions

All these categories are defined by one main phenomenon and that is the notion of risk. There are different types of risk within road safety that range from intentional risk taking to unintentional risk taking, all of which play a part in determining certain types of accidents and delineating the type of people that cause them in certain spatial locations. This section discuses the main high risk road user types which are dominant in the literature. These include risk groups by ages, particularly children, teenagers and the elderly, and by gender. Secondly, by user group, such as pedestrian, cyclists and motorcyclists thirdly, the risk of non natives interacting with the road environment.

There has been a wealth of literature which discusses the relative differences that age poses on collision involvement. For example see Abdel-Aty *et al* (1998), Claret *et al* (2003) and Oxley et al (2005). Many of these research papers address all ages and possibly in different circumstances (for example pedestrians or specifically drivers).

Older drivers have been the subject of much debate, particularly when the UK, Europe and America's population is ageing (Bort 1994, Sjogren 1996, Dulisse 1997, Keskinen *et al* 1998, Zhang *et al* 2002, Baker *et al* 2003 and Hing *et al* 2003). There will be a need in society to address the needs of the increased numbers of older drivers and to understand what the implications are of this increasing age group who are on the roads. Dulisse (1997) points out that the physical skills of driving erode with age, however usually as a consequence older drivers take more care and attention whilst driving. Research by Keskinen *et al* (1998) concentrates on the potentially dangerous interaction between older and younger drivers at junctions, indicating the likelihood of older drivers to suffer more severe injuries. The innovative research by Baker *et al*

2003 suggests that road user risk is becoming polarised and prospect of an ageing population within road safety has not been properly addressed.

The elevated collision risk of teenage drivers has been well documented These teenage drivers are aged between approximately 16-19 with their younger counterparts being at risk from being involved in a collision as a pedestrian and cyclists. However, the first, older teen risk will be dealt with first. A number of studies have researched this elevated teenage risk (Ulleberg 2002, Williams 2003, Dissanayake 2003).

There is a certain presumption that there is a predetermined risk group, being young male drivers driving in the evening and late at night. A study in Denmark sought to ask the question; Is there a risk taking group of fatigued young male drivers, driving at night? The study used a roadside survey method and concluded that there is the presence of a risk taking group who are however aware of the risks they are taking, having the potential to fall asleep and are ready to accept it (Corfitsen 1999). Comparing age groups in terms of road user risk can be misleading. Depending on your age and gender people have varying degrees of mobility and different interaction with the road environment. Children for example are more likely to be a pedestrian or cyclist and their risk will be higher because they are more inexperienced and vulnerable. Analysis by Petch *et al* (2000) concluded that exposure was a strong indicator of road collision risk, however there was no simple explanation for the spatial distribution of child/cyclist casualties at a district level (in Salford near Manchester).

Gender has always played a strong dominating factor in determining someone's risk of being involved in a road collision. Generally, in the London dataset there is a higher proportion of collisions involving male drivers. This could be partly due to the higher road use mobility of males within the road environment rather than females, however this is just conjecture. Although in western countries, such as the UK the number of female drivers has increased (Laapotti *et al* 2004) and the difference between the number of males and females involved in collisions is decreasing. However results have been confusing. McKenna *et al* (1998) compared the collisions from females and males between 1979 and 1997. McKenna *et al* (1998) concluded that the sex difference in the pattern of collision involvement is changing over the years. This is inconclusive and therefore means more research is necessary to determine the differences between male and female collisions.

In the earlier section it was mentioned that there has been little direct evidence on the relative road safety of immigrants. Recent American studies have identified race as an important road safety issue. A study by Dobson *et al* (2004) focusing on the increased collision rate among the immigrant population in New South Wales broke down the road users into drivers, passengers, pedestrians and other road users. However the results were inconclusive, they indicated there was no evidence to suggest that drivers born in other countries were more likely than Australian born drivers to be involved in collisions resulting in death or injury requiring hospitalisation. As with the majority of literature in this area of determining road user risk groups, there was little or no spatial dimension which looked at the neighbourhood variations at a local level. It is useful to be aware of this literature for the purpose of this typology.

Road user groups are often distinguished by mode of travel, in this case, cyclist, pedestrian and motorcyclists. All are identified in 'Tomorrow's roads: Safer for Everyone' as high risk road users. Research by Noland *et al* (2000) into pedestrian and bicycle casualties in the UK has indicated that per capita income has been found to be positively associated with road collisions (this was negative for pedestrians). The study also found that there was a strong association between bicycle collisions and the age group 0-14, but this was not significant for pedestrians within the same age group. There has been in recent year's concern for the risk of older motorcyclist users in the UK. Although the research into their risk has been limited, they are still recognised in collision statistics as a risk group. Predominantly these riders are inexperienced and there are a high proportion of collisions which occur on bends in the road (Carr *et al* 1995).

2.4 Speed analysis

Speeding is a common cause for a high percentage of road collisions in the UK. Any thorough research of road collision causation would include speeding as it has been proven by many academics and analysts to play a large role in the propensity of road collisions. Speeding can be seen from two different view points. The first is that speeding is inherently a human behavioural phenomenon. The second view point rests on an environmental explanation that driver's speed due to road surface, road design and surrounding land use. For example a driver is more likely to speed on a rural straight road than in a dense built area with bendy roads. This research and review is going to focus on speeding as a human behavioural issue because the inclusion of geodemographics and socio economics makes this view point more feasible to analyse and explain.

The laws of physics tell us that force is a product of mass and velocity combined therefore the greater the speed the more severe the collision is likely to be. Therefore the first question to answer is why drivers speed. In a survey by Brake (UK road safety charity - 2002) found the top three reasons for people speeding were:

1. Just not realizing
2. They were late/in a hurry
3. They forgot or did not know

The general perception of drivers is that there is little risk involved in speeding from the viewpoint of not being caught and being involved in a collision. In the same survey it was determined that more male drivers speed than female drivers (the ratio was 5:1). Within the male grouping it was disaggregated and found that there was a high propensity of young male drivers who ignored speed limits. A study by McKenna *et al* (1998) analysed the impact of passengers on the drivers speeding habits and they found that an absence of a passenger resulted in males driving faster than females. Where they are male passengers in the vehicle it resulted in faster speeds and when females were passengers in the vehicles it resulted in slower speeds.

It has been well documented that faster traffic (and therefore speeds) results in more collisions (see Taylor *et al* 2000), as well as the notion that the larger the spread of speeds around the average increased the number of collisions. A study by the Department for Transport, Local Government and the Regions (DTLR) in 2000 analysed the types of roads in which speeding was more likely to occur, the results illustrated that 69% of drivers speed on urban 30 m/p/h roads, whereas 55% of drivers speed on motorways. This highlights the strong need for a better understanding in the patterns and trends of speeders in London as the majority of roads are 30 m/p/h in built up areas. This is supported by Noland *et al* (2004) who determined that 82% of fatal road collisions in London occur in 'congested periods' when actually the traffic speeds are lower than average.

A road collision may occur due to variables at several different hierarchical levels. These include

Regional
Area Wide
Road segments

Independent Variables (pedestrian location/movement)
Individual (Mobile phone use, attitude)

The risk of speeding can occur at all these different scales and will have different effects accordingly. Speeding in the London and the UK is perceived by the police and local authorities as a crime and essentially this type of crime can be prevented which has been seen by introductions such as the Safer Speeds Initiative however this will be discussed further in the management section.

2.5 Risk factors – Environmental

2.5.1 Weather conditions

Edwards (1996) suggests that although weather may not be a principal component in collision causation it does play an important role. Several studies have demonstrated a relationship between weather and road transport (for example see Andrey and Olley 1990). It could be assumed that road collision risk increases during adverse weather conditions however it is important to note that there is no simple cause and effect relationship between the two. It has been clearly demonstrated that during wet conditions (i.e. rain) collision numbers increase (see Brodsky and Hakkert 1988) partly due to the numbers of people who would take the car as opposed to public transport or walking. Another reason for the increased propensity of collisions during wet weather can be explained with reference to Adams work on seat belt risk (Adams 1995). Increased safety devices in vehicles such as anti locking brakes means a possible increase in risk taking. Adams supports this by citing other safety features such as seat belts increasing the propensity of risk taking amongst drivers (Adams 1995).

Edwards (1996) undertakes a county wide investigation with a crude demonstration of variations in weather related road collisions from the STATS19 police database. The study rests on the evaluation of the coded weather variable for each collision (see appendix for all STATS19 variable codes) whereby county rates were compared. The research determined that the majority of local authorities had between 75-80% of all road collisions occurring in band 1 – fine weather with no high winds. The study is limited to only analysing at a county level and relying solely on the police data which Edwards mentions is under representative of the true picture of a large proportion of road collisions. The study based in the UK highlights the need to supplement the data with meteorological data such as precipitation rates or number of hours of sunlight each day to ascertain any comprehensive and reliable conclusion (Edwards 1996).

A more comprehensive study focusing on the effects of rain in a range of cities in the USA by Eisenberg found contradictory evidence to suggest that the presence of precipitation actually reduced the number of road collisions (Eisenberg 2004). In contrast his work found that drivers are at an elevated risk when precipitation falls after a dry period for several days or more. This research suggests a more detailed explanation of the influence of weather namely precipitation on the risk of being a factor in road collisions. Weather clearly plays a role within the risk of a collision and this risk needs to be understood in terms of asking questions such as what sort of person is more likely to be involved in a collision when there is adverse weather and where are these collisions more likely to happen to these people? The research seeks to get a better understanding of the role of weather in road collisions and what its relationship is with other collision variables such as time.

2.5.2 Temporal analysis

There is a considerable amount of literature on time geography (for examples see Chapin, F.S 1974, Carlstein *et al* 1978 and Friedman M 1983) however its application to understand the temporal patterns of road collisions have been limited. Time – space geography had its origins from Torsten Hagerstrand's 'space time model'. Time geography studies the space-time behaviour of human individuals; in their daily life people follow a space time trajectory.

The time space model is an attempt to understand under what basic condition linkages like collisions develop. Transportation and therefore road collisions are fundamentally space-time orientated, where the human population is conceived as forming a web of paths which flow through a set of time space locations (Carlstein *et al* 1978). Each path, or for the sake of this research we shall call journey has a life span (a journey time), however these paths are not isolated, in terms of transportation they co exist along the road network. The temporal importance of road collisions is very prominent within the factors that determine the risk of a collision. The movement that drivers take through space or in other words from A to B takes up time and yet has constraints surrounding it for example attitude to speeding, type of vehicle and weather. These factors all contribute the changing temporal dimension within a road collision.

As mentioned time geography research for road collisions has been limited. The predominant focus has always been larger temporal scales such as yearly or monthly analyses. This type of analysis is often used for entire countries or cities. Levine *et al* (1995) postulates that there have been no studies which documented the day to day changes in road collisions for an entire metropolitan area. The primary reason for this larger scale has been the limited number of road collision cases to analyse on an hourly or daily scale, with the combined challenge of the differing time scales which make comparisons difficult. A study by Folkard (1997) introduces the notion of 'black times' in a 24 hour day. The methodology uses Z scores and Poisson Regression to analyse the daily temporal patterns. The fifteen year dataset for Seattle USA highlighted collision risk to be at its highest in the early hours of the morning (between 3am – 7am). The second peak was found to be in early afternoon. His research draws upon examples from both maritime and industrial temporal studies particularly focusing on patterns of human sleep, work and eating. By understanding the diurnal habits of humans and taking into account the high risk peaks at 3am and 3pm, these time periods were found to be high risk periods in industrial studies for workers. It has been suggested that the individual's circadian rhythms are the cause of these peaks. However Folkard (1997) argues that these circadian rhythms are insufficient to account for the variations over a 24 hour day.

A different approach and more worthy of evaluation is work by Levine *et al* (1995). The research (using data from Hawaii) looked at ten variables from the collision data to analyse their relationship with the daily time periods using a regression framework. The variables included whether there was a 'major public holiday' or a 'minor public holiday', day of week, weather, unemployment and tourist population rates. Results showed a strong cyclical variation in daily collisions primarily due to the strong periodicity in daily traffic volume. However the variables that did show significant coefficients were traffic volume, Fridays and Saturdays, minor holidays and rainfall. All these factors influenced the daily temporal patterns of road collisions. This methodology and some of the variables will be applied to the research methodology, such as public holidays and traffic flow. Rainfall will also be added as weather as a factor in road collisions has been subject to much debate and exactly how much influence it plays in the chain of events that increases the risk of a collision.

2.6 Infrastructure and road safety

Interactions between land use and transportation decisions play a huge role in a person's propensity to be involved in a collision depending on where they live. Many studies have explored the combined effect of roadway geometries and environmental factors on road safety. The spatial environment can be understood in many ways. For example land use was considered by Petch *et al* (2000), Ivan *et al* (2000), Ossebbruggen *et al* (2001) and Noland. Land use, infrastructure and transportation networks play a significant role in determining road user risk. Its changing role and dynamic within a city has been discussed in detail by Batty and Longley (1994). The growth of a city outwards will almost never be exactly concentric and even, cities usually organised into neighbourhoods enough to support educational and retail functions (Batty and Longley 1994). These patterns in turn will affect the propensity and location of road collisions. The significant environmental and spatial factors which relate to the changing city attributes with distance from a city centre include changing land use and changing road network (usage and density) (see Anas *et al* 1998). London's land use and infrastructure is unique with respect that it is a capital city and its growth has produced an agglomeration whose road network, land use and city centre have experience continued growth and change.

There has been a strong interest in the relationship between road collisions and the characteristics of roads and local environmental conditions. There have been many studies which have examined the relations between vehicle collisions and road way geometrics (Agent *et al* 1975, Zeeger *et al* 1990, Ivan *et al* 1999 and Martin 2002). What is being considered here however is the nature of the adjacent environment of the roads, this can be analysed in a number of ways. For example land use was considered by Ivan *et al* (2000) and Ossenbruggen *et al* (2001). Hamerslag *et al* (1982) studied the location of bus stops, parking bays etc. Whereas Henning-Hager (1986) examined the relationship between residential development and road safety and Abdalla *et al.* (1997) researched the relationship between road collisions and the effects of areal characteristics.

2.6.1 The urban landscape and road safety

Daily trips in an urban environment will take place over an ever expanding perimeter covering several neighbourhoods and even small towns (Millot 2004). Interactions define the urban area, with this interaction and traffic function of urban roads, there is often a clash of interest. In the literature there has been little investigation into the effects of urban growth and road safety.

2.6.2 Land use planning risks

Planning decisions regarding land use and the road network have a significant impact on the volume of road collision. Without strategic land use planning practices, residential, commercial and industrial land use will evolve ad hoc as will the road network to meet its needs. The effects of this lack of planning would result in inappropriate interaction between road users and land use. For example there would be increased volumes and speeding traffic through residential areas increasing the risk of road collisions, due to the increased potential for interaction with pedestrians.

2.6.3 Public amenities, shopping centres and collision risk

Urban areas since the 1970s have been experiencing a movement or migration of residents from the inner districts to the suburbs and London is no exception. Socio economic changes have bought about an increase in out of town shopping centres and shopping 'belts' for example Brent Cross and Bluewater have lead to a decrease in local shops and the notion of the 'high street'. This in turn has increased traffic volume on roads surrounding these shopping centres, with the predominant form of transport to get there being the private car. This increases the risk of collisions due to the high volume of traffic and minimum public transport.

2.6.4 Road related risk factors

Road collisions tend not to be evenly distributed along the road network. They tend to occur in three main spatial patterns:

1. Clusters at single sites (for example a roundabout or junction)
2. Along sections of roads
3. Scattered across whole residential areas (especially areas of social deprivation)

The road network plays an important role of collision risk because its layout determines how road users perceive their environment and also provides information for how to use it safely through traffic signals and controls. In the planning, design and maintenance of the road network Ross *et al* (1991) outlines elements effecting road safety:

- safety awareness in the planning of new road networks
- the incorporation of safety features in the design of new roads
- safety improvements to existing roads
- remedial action at high risk incident sites

The WHO report on road safety (2004) outlines specific situations related to road planning that are risk factors of incidents:

- Through traffic passing through residential areas
- Conflicts between pedestrians and vehicles near schools located on busy roads
- Lack of segregation of pedestrians and high speed traffic
- Lack of median barriers to prevent dangerous overtaking in single carriage ways
- Lack of barriers to prevent pedestrian access onto high speed dual carriageway roads

2.6.5 Management of risk through planning and land use intervention

Rumar (1999) has summarized that the following strategies can reduce exposure to road incident risk in a planning and land use environment focusing on the road network:

- reducing the volume of traffic flow of motor vehicle traffic by means of better land use
- providing efficient networks where the shortest or quickest routes coincide with the shortest
- encouraging people to switch from higher risk to lower risk modes of transport
- placing restrictions on motor vehicle users within the road infrastructure

2.6.6 Effective use of land use

The organization of land use affects the number of journeys people need to take, means of transport, length of trip and route taken. In short different land use creates a different set of traffic patterns. Hummel (2001) outlines the main aspects of land use that influence road safety are:

- the spatial distribution of origins and destinations of real journeys
- urban population and density in pattern of urban growth
- the configuration of the road network
- the size of residential areas
- alternatives to private motorized transport

Hummel (2001) suggests that land use planning practices and 'smart growth' land use policies and coupled with the development of high density, compact buildings with easily accessible services and amenities can lessen the risk exposure to road users. The creation of clustered mixed use community services for example can cut the distances between commonly used destinations, cutting the need to travel and reducing dependence on private motor vehicles.

2.6.7 Planning for safety awareness

Road safety considerations are central to planning, design and operation of the road network. Altering the design and layout of the road and the road network to accommodate human characteristics and human error road safety engineering can be successful in injury prevention and reduction. Ross *et al* (1991) outlines a framework for the systematic management of road safety which is outlined below:

- classifying the road network according to its primary function
- setting appropriate speed limits according to this function
- improving road layout and design to encourage better usage

2.7 London Government and borough policy

London's Road Safety Plan is the framework which provides London with road collision reduction strategies and targets (Transport for London, 2004). It covers all of Inner and Outer London as one entity and highlights a threefold need to:

1. Concentrate on safety through partnership
2. Manage speeds – by reducing excessive and inappropriate speeds
3. Protect vulnerable road users

These aims primarily use past statistical evidence of collision events. This STATS19 database collects over fifty variables for every collisions including information on gender, age and what is being used for this study, postcode data. This London plan is loosely based on the framework put forward in the UK wide road safety proposal 'Tomorrow's Roads: safer for everyone' (Department of Transport 2001).

The London road safety plan identifies patterns and trends of road collisions occurring to a wide spectrum of road users and also vehicle types. However the main focus of analysis of Transport for London is understanding patterns of road collisions occurrences with particular reference to vulnerable road users (for example, the elderly and children) and how their involvement in collisions varies spatially (see LAAU topic 2001-6, 2001and LAAU topic 2001-1, 2001, Levels of Accident Risk in Greater London, Issues 9,10, 2003 and 2004). It is evident from this literature that a broad spatial understanding of collision propensity for all London's residents and how this propensity varies spatially is being neglected.

This broad spatial understanding mentioned above can be subdivided into two main challenges facing the road safety management in London primarily at the borough level. At the latter level, each borough has individual policies for road safety management and reduction. One of the consequences of this is the restriction of not being able to compare outcomes, schemes and residential road user risk patterns across boroughs accurately. The second challenge faced is at a 'local' level meaning smaller than a borough level whereby as with the challenges at the borough level there is no spatial framework within which analysis can take place and be compared with regards to residential road user risk and how this changes with distance from central London, which is a notion which can be applied to all boroughs.

Approaches to road safety vary from borough to borough. However key themes are notably London wide, for example education initiatives for vulnerable road users such as children. Although they may not be analytically similar in terms of say for example, variables collected in order to analyse the risk the aims of reducing collisions are universal. With regards to analysing road user risk with relation to where they live and their socio economic status, it is a potential tool for road safety policy makers. The analysis of postcode data of the drivers and casualties living in a particular borough (whereby their collisions may not necessarily occur in the same borough) would potentially lead to a deeper understanding of residential risk patterns (and eventually the different types of collision occurring, for example the level of severity or whether they are more likely to be involved as a driver or casualty).

It is evident therefore that there is a necessity for the analysis and understanding of risk for not just what the data identifies as 'vulnerable' road users but society as a whole. Coupled with this analysis there needs to be quantitative measurement of the differing collision risk levels people experience which may change for example over time or depending on the type of journey they make.

The London Accident Analysis Unit (LAAU) from July 2000 became part of the London Road Safety Unit part of Transport for London Street Management. This Unit provides collision research for London in order to provide policy recommendations. In 2003, the LAAU published information on levels of risk. It uses percentage levels of various types of collision, and collision rates per site per year are provided for certain roadway features such as roundabouts, and pedestrian crossings. Although this gives a sound impression of the pattern of collision risk, there is limited appreciation for spatial and temporal patterns. The collision rates are displayed as rates per site or kilometre of road. The disaggregation of risk has been ascertained for every collision within a 20 metre radius of an automated traffic signal junction by borough. This is predetermining collision hotspots based a potential high risk location of traffic lights. This could potentially give an inaccurate understanding that the reason for the high number of collisions is due to the spatial location near the traffic lights. However, this could mask the underlying reasons for the collisions to occur, which in itself is subject to a high degree of error. This is due to trying to recreate the collision scenario based on historical data and circumstantial evidence.

There has been call in recent years for the management of road collision reduction to be treated as a crime such as burglary or murder by government officials. Transport for London estimates that 90 road deaths a year occur due to speeding (190 people are murdered in London). Jones (2003) argues that the Metropolitan Police should play a larger more pivotal role in reducing speeding and argues that the reduction in traffic police have not helped to reduce the number of fatalities on London's roads. The key strategy from the research by Jones (2003) and the London Road Safety Plan (2001) is the greater need for partnership amongst road safety professionals particularly across boroughs and police forces.

2.7.1 Current road policing policy in London

Road traffic policing across London provides a reactive as well as proactive service for the prevention and reduction of road collisions. The police respond to road collisions, collect data and assimilate reactive policies concerning reduction. In terms of road policing, the approach is centralised to reducing 'road crime' which may or may not result in road collisions. In many police surveys the public have rated safer roads on a par with low crime rates and therefore increasingly crime and disorder strategies include a road policing element (HMIC Road Policing Traffic Report 1998). Current research has been working on profiling neighbourhoods in terms of geodemographic classifications and how people rate the police and how they feel about crime and this theoretically can be applied to the theme of road safety which as mentioned is inter linked with crime reduction.

The HMIC report identifies many police forces managing road-policing intelligence on a 'piecemeal, ad hoc basis which is often personality driven' (HMIC report on road policing 1998). The idea of partnerships also features heavily in the report and states it is important to distinguish between a partnership and a liaison with other agencies. Partnerships involve two or more agencies jointly designing a common strategy, monitored by or on behalf of all parties and often involving the pooling of resources. On the other hand, multi agency liaison involves several bodies, each with its own aims and strategies co-operating with each other because it is beneficial to work together.

A study concerning 'Traffic Policing in changing times' was conducted over 10 years ago and there have been no subsequent studies since which seems to suggest that the role of policing and specifically the spatial aspect as received very little attention. The main principle to be acknowledged with regard to road policing is that the police cannot control or reduce every collision because the cause of the most collisions will be beyond their jurisdiction. Road collisions are a product of a range of interlinking risk factors which present themselves at certain times and locations. The police are only able to manage a small number of these collision risks for example speed control and dangerous driving.

In response the research which outlined that policing needed to be either 'policing by objectives' or using 'traffic policing priorities' has seemed to be applied in many of the police forces in England and Wales over the last ten years. Policing by objectives can be defined as 'setting goals and objectives, developing action plans which provide guidelines for control and corrective action' (Lubans and Edgar 1979). As mentioned earlier police force goals vary from force to force but there should be some universal guidelines for education, training, supervision and communication. The research outlined the amount of time spent on different traffic policing activities. This only supported what academics and police officers already knew which was that the majority of their time was spent completing paperwork. There are so many contentious issues surrounding these activities such as whether the police are stopping the right people, and whether that police action has any significance impact upon driver's subsequent driving behaviour.

Many of the solutions to traffic safety lie in things such as road engineering which is out of their immediate control. The impact of this kind of police activity on road collision figures is extremely difficult to measure. In the concluding remarks of the study it outlines the options for traffic policing problems, which consist of:

- Police officers needing to be more aware of the wider economic, social, political and environmental issues concerning management
- Physical preventative measure, i.e. that the roads need to be physically safe

- Technology
- Extra police man power

This is only one interpretation to a very complex problem encompasses many issues.

The role of this research within the scope of road policing is a multi faceted one encompassing many areas of road safety the police are responsible for. By identifying a more comprehensive understanding of the drivers and casualties involved in collisions within London society, the police will be able to possibly combine this information with crime reduction; it also means the education and policing strategies will be more intrinsically targeted to people and areas that are most at risk from certain types of collisions. By determining a spatial dimension to the research, the proactive responses to collision hotspots will be better allocated.

2.8 Collision hotspots - critique of methods

Much of the literature surrounding road collision data focuses on prediction and evaluation. There has been little work on the subject of hotspot collision definitions within London and indeed the UK. In recent years there has been a renewed interest in this delineation of collision hotspots due to the awareness of spatial interaction existing between contiguous collision locations, and therefore suggesting a spatial dependence between individual occurrences (Flahaut *et al* 2003) This section seeks to address the challenges of this deficiency of clarity surrounding the question of how to quantify and define a collision hotspot. Studies in the field of road safety focus on three different types of collision cluster which has been identified as:

- LINK (road segments)
- NODES (junctions or roundabouts)
- CELLS (area wide such as a housing estate or car park)

The rationale supporting the evaluation of different types and explanations of collision hotspots rests on the main theme central to this research which is to classify hotspots within London relating to a wide range of engineering, socio economic and enforcement variables. This section will be subdivided into two sections, the first focusing on the definitions and measureable collision variables that are used both in London and the rest of the UK. The second section highlights the different methods associated with identifying the hotspots (using methods and examples from the UK, Europe and New Zealand).

2.8.1 Quantitative and spatial methods of road collision analysis: a critical review

The methods to identify patterns of road collision occurrence have really evolved in the past twenty years. Modelling has become one of the first and foremost applications in road safety analysis (Smeed 1968, Smeed *et al* 1970, Mekky 1985, Gharaybeh 1991) predominantly because of its universal appeal to predict and forecast levels of road collisions. However in the earlier years the analysis was not comprehensive and ignored the impact of variables which may contribute to a road collision. These early models used multiple linear regression modelling with its assumption of normally distributed errors, but there has been a steady realisation that the nature and occurrence of road collisions is such that it is far better to model the process using the assumption of a Poisson distribution for the frequency of collisions in a given period of time at any one site (Maher *et al* 1996). Since these early prediction models, research has been focused on incorporating causal factors into the prediction models such as traffic flow. There has also been considerable research to develop models to predict the number of collisions for individual elements of the urban transportation network such as at pedestrian crossings (Tarko *et al* 1995) or minor junctions (Mountain *et al* 1996). However these models have a number of drawbacks which have been outlined by Leveson (2004). These include the need for a better understanding of the relationship between humans and automation. This had lead to a need for understanding human behaviour, their choices and error when interacting in the road environment. In addition to this, there is a need for a better understanding of the complexity that faces the road environment and risk of a collision. Complexity has many different facets, most of which are increasing in the systems we are building, particularly interactive complexity (Leveson 2004). Another major drawback which has been identified by Thomas (1996) is that the large majority of the clustering models rely on using road segments of a specific length, and in a small area. There seems a lack within the literature of discussion of the optimum segment length and the scale of the studies. This important aspect of modelling has been addressed by Thomas (1996) with regards to network autocorrelation.

The analysis of road collisions over the past ten years has evolved with regards to increasing cost benefit needs, this has meant a shift to analyse and prioritise road collision hotspots using a variety of methods. Road collisions were seen as spatially dependant and their tendency to cluster in certain areas had to be understood. Collision reduction was loosely based on four types: single site, route action, area action and mass action plans (Nicholson 1999). The identification of hotspots became a priority and more importantly the reasons for why they were occurring. These methods therefore to identify collision concentrations traditionally apply to identifying hotspots

(Silcock and Smith 1985, Nguyen 1991, Joly *et al* 1992 and Hauer 1997). Since the 1990s there has been a strong increase in the potential of using GIS to identify these hotspots and suggest reasons for causal dependence. The most straightforward use of GIS for collision analysis is the examination of spatial characteristics of collision locations (Steenberghen *et al* 2004). The use of GIS and spatial statistics has propelled the study of collisions, and instead of merely using a visual interpretation of the collisions, GIS and its associated spatial statistics has meant that researchers have been able to understand the spatial dependence with statistical significance. Methods such as point on line overlays and point in polygon overlays make it possible to add attributes of the road infrastructure. Spatial clusters can be further used for pattern detection in a GIS environment (Openshaw 1995, Openshaw and Turton 2001). Two dimensional clustering in GIS uses kernel density smoothing and determines spatial clusters using density of the points. This method has gained momentum in recent years due to its accessibility, statistical significance and ability to manipulate the method. Studies by Flahaut *et al* (2004), Steenberghen, *et al* (2004) and Sabel *et al* (2006) have indicated significant results for using this method for road collision analysis. This method allows the user to differentiate between route and area hotspots within the dataset. The cluster detection methods in all these studies are grid based and a grid cell is used as the basic spatial unit. Traffic safety is ultimately the result of a balance between the type of traffic and road and neighbourhood characteristics. Using GIS bridges the gap between statistical analysis and visualisation in order to effectively analyse road collisions and their causes.

2.8.2 Definitions and quantifying collision hotspots

Within the UK (with exception of London) road safety and collision management are handled by a range of public bodies, namely the local authorities and the police authorities which conveniently share the same administrative boundaries. London's road safety is managed by Transport for London and the Metropolitan Police. At a more disaggregated scale Local Borough Authorities and Police Authorities manage the detection, management and evaluation. It is important to acknowledge that the collision hotspot approach to collision reduction represents only one of a number of low cost engineering strategies in the road safety field.

It is clear from the early literature (for example Silcock and Smyth 1985) that the importance of route, area and mass plans were predominant. The scaling of hotspots is also relevant in this section, for as the locations with the heaviest concentrations of collisions get identified there needs to be a ranking system in order to prioritise the importance of the hotspot for resources and attention. This theme will be discussed in more depth in Chapter Three and Five respectively. This ranking must also take into account on which type of road network it occurs on, as mentioned earlier either link, node or cell. The rankings can be based on many different variables such as severity, collision count, collision rates, types of collision (such as child pedestrians or elderly) etc.

2.8.3 Usefulness of different methods

By evaluating spatial collision hotspot methodologies it is clear that disparities do exist between them. Attention must be focused on the requirements for the research in order to determine the most suitable method. Overall, the research seeks to collate the factors in each cluster determine over various temporal measurements (hours, days, weeks) to determine and understand the changing influencing factors on collision risk. Therefore a method such as kernel density estimation will be used for the initial detection of the collision hotspots for analysis. This method provides the initial advantages of taking absolute numbers of collisions with no interaction with traffic flow so as to identify 'static' collision prone locations. It can initially provide scope for manipulating band widths and parameters and within those different identified hotspots further analysis can be carried out to ascertain the influencing variables. Following the visualization of collision hotspots, multiple regressions are intended to be carried out on hotspots over different time periods.

2.9 Research gaps and thesis aims

Referring back to the end of Chapter One this section seeks to formalise the link between the literatures reviewed in this chapter together with the hypotheses and aims for the thesis. This sections aims to outline the gaps within collision research and how best this thesis can approach these gaps and the intended outcomes.

In the late 1980's Whitelegg's (1987) influential article about the geography of road collisions he claimed that geographers do not pay enough attention to road collisions. It is true to say, that this statement can still be applied today. The reason for this argument is that, the analysis of road collisions has been continued to be neglected by geographers and instead taken over by engineers, statisticians, planners and behaviourists. Inherently therefore the geography of road collisions has been neglected for the last twenty five years. It has been only the last ten years that we have seen an 'explosion' of road collision research especially in the areas mentioned previously and of course, using GIS. However I believe there has been a leap of research from this neglect of geography to a prominent use of GIS to explain road collisions. This leap in research has meant that Whitelegg's (1987) claims went without reaction until the revolution of GIS in the mid to late 1990s which saw an out pouring of research in this area. It is my argument that there needs to be a 'step back' to the actual geography of road collisions in order to understand the complexity of road collisions in a spatial framework. There also needs to be better understanding of how all the other dimensions of collision analysis (statistical, engineering, and behavioural) can be incorporated into the geography. The geography therefore should be seen as the framework within which to study road collisions, however this has not been the case in much of the previous collisions research. This thesis aims to address the neglect of geographical framework by examining the spatial nature of collisions and addressing the issues and limitations of using geography, but also advancing our knowledge by using up to date analysis tools.

In the literature on road collisions and spatial data analysis there is little discussion or apparent appreciation of the collision data and its limitations and drawbacks associated with road collision analysis. This neglect applies not only to the spatial data but also the analysis of it and there is minimal attention to this in the literature which reinforces this need to analyse collisions in a geographical framework. In response to this deficit in the literature, this thesis seeks to address these geographical limitations, not only critiquing the data but also the techniques of analysis. It will assess for example the drawbacks of aggregated analysis, uncertainty and accuracy. These are all important geospatial data issues but ones that have not comprehensively been discussed in the context of road collision analysis. Chapter Three investigates the nature of the road collision data, collection and reporting processes, followed by a robust description of the data analysis limitations, such as spatial/network autocorrelation, scale and ranking of hotspot locations. As mentioned, this is a unique facet to the thesis, as these spatial limitations of road collision analysis are neglected in much of the literature.

There are many areas of road collision research which focuses specifically and separately on risk and socio economics in relation to road collisions. No attempt has been made to conceptualise the notion of risk and quantify it with regards to socio economics (related to the casualties and drivers). This thesis attempts to bridge this divide between road collisions, risk and socio economics using a variety of methods and analysis tools. Therefore one of the aims of the thesis is to identify 'risk factors' based on the road collision location and the environment from which the casualties and drivers reside. Traditionally, the analysis of risk factors has focused on a small sample of attributes relating to the road collisions. This is some ways can be perceived as a disadvantage and why in this thesis there is an inclusion of a large number of variables. This inclusion of a large number of variables expands the search to find influencing factors which contribute to collision occurrence. As collisions are a comprehensive event which are made up of a number of interacting attributes rather than just one explaining variable.

A large proportion of the literature identifies what I have called 'high risk' groups within this chapter in order to explain groups within society who have a higher likelihood of being involved in a collision than other people. Often these high risk groups are aggregated by gender or age or behavioural type but there is an inherent lack of spatial understanding into the occurrence of these groups. This thesis aims to try and bridge this understanding by adding in the spatial dimension to high risk groups within society. It is straightforward to identify who the high risk groups are but more challenging to add a location to these groups which is what this thesis attempts to do. Similarly, the temporal element of collisions has been somewhat neglected over time, supported by Levine's (1995) comment that there have been no studies which have documented the day to day changes in road collisions for an entire metropolitan area. Although time is not the predominant variable in this study, it plays an important indicator in the hotspot clustering process, by ultimately indicating the most prolific times for each hotspot in the overall typology.

From this outline of research gaps it is possible to outline some statements which represent the some general aims of this thesis before going on to discuss the nature of the aims in finer detail. Overall, I believe there is a need within the literature which addresses the nature of risk in relation to a person's socio economic characteristics and road collision propensity. Previous studies as mentioned have lacked this quantification of risk. In short, this thesis aims to set a broad measure in statistical terms with which to understand a person's risk. Collision risk in transport is often attached to very limited factors concerning vehicles, road systems and driving behaviour among individuals. However regional variation could be due to broader collision patterns due to different physical structural factors, socio economic and cultural differences. It is this last sentence which bears the most applicability to this thesis and one that needs to be built on. In the same way it can be argued that road collision risk is closely connected to risk conditions in other areas of people's lives. Although this is inherently problematic to prove, the socio economic and lifestyle indicators give broad indication with relation to financial risk and lifestyle components.

Therefore, one of the backbone aims for this thesis is to create a GIS based road collision risk model of involvement in the collisions. By understanding people's involvement in collision and ultimately their risk, it has the potential to improve the public perception of road collision risk. This ultimately could mean challenging lay beliefs of people with regard to their risk. This thesis aims to undertake a broad risk road collision risk assessment requiring both an understanding both of the people involved (defined by the environment in which they live) and the physical environment of the collision location. A GIS platform is especially suited to this type of problem because it provides an efficient system of linking a large number of disparate databases. It also provides a spatial referencing system for reporting the output at different levels of aggregation. The thesis is disaggregated into two areas of aims, split by the locational aspects of a collision. Questions concerning the collision location are broadly placed around the need to find an effective tool for grouping collisions together using GIS. This method focuses on 'effective' and consists of a number of rules by which the grouping method has to adhere to. For example, the criteria would include the following:

- Not include the road network
- Use a 'basic spatial unit' as the building block for the road collision hotspot
- It would be a measure of density based on the absolute number of collisions
- It would be able to incorporate information regarding the temporal and casualty and driver information

There is a need therefore for a more systematic approach to the identification of road collision hotspots which can automatically detect statistically significant spatial collision clusters offering repeatable results by removing the subjectivity.

The second area refers to the residential location of the casualty and driver. The underlying notion of this rests on the theme that people of different (varying) socio economic status have different levels and types of collision risk in London. This assumption is built from a number of smaller assumptions which make up this section of the research. Namely that when assessing an individual's risk there is an element of 'cultural filtering' (Adams 1995) which affects the risk and outcome of different people according to their cultural status and background. By identifying that there is a difference in collision risk between people, there is a need to understand why these 'high risk' groups have a higher propensity to be involved in a collision and how does this vary spatially and temporarily. Although I have not mentioned geodemographics in this section, because it will be reviewed in depth in the next chapter, where there will be an overarching theme of whether geodemographics provides a robust framework within which to study a person's risk?

Ultimately, can their environment (where they live) effect if/why/where they have a road collision?

2.10 Literature review: a summary

In developing strategies to determine road collision hotspots and their causes for both an environment and socio economic point of view, a number of research domains were discussed in order to cohere together the main points in order to understand and develop a concise research process. The aim of this chapter was to give a more detailed conception of the literature and how relevant this research is to the road safety literature as a whole. However it is worth noting that a more succinct review of literature relevant to each chapter will be found at the beginning of each chapter.

In the attempts to delineate new high risk road user groups there is substantial literature, which goes further than using the traditional vulnerable groups within society. Research by for example, Pietro (2001) and McKenna *et al* (1998) has increased awareness that the risks drivers and casualties make in the road environment may be linked to a larger scale of influences, including gender, social and family patterns, housing, age and attitude. We know that children from disadvantaged areas are more likely to be involved in a collision (Scottish Executive 1999); however we are unclear as to this relationship with adults, who use the road environment on a more frequent basis. In the same way we know that children from ethnic minorities are more likely to be involved in a road collision (Christie 1995), however, again, the relationship is unclear for adults. The literature therefore highlights many questions that are still unanswered and that this thesis can help to answer.

So far in road collision literature there has been no direct study linking hotspots to its victims and their characteristics. This research therefore aims to determine the most vulnerable road victims based on the collision density they are likely to be involved in. Defining both of these (the hotspot and the risk) has been subject to many different methods which all have their strengths and weaknesses and have led to the overarching methodology in this thesis. There is an overall pattern in the literature of trying to understand the nature of risk which contributes to a road collision. It is important however to utilise the data from the collisions which have a spatial dependence and therefore could constitute a hotspot. The people involved in these collisions share a commonality, in so far as they were all involved in a collision in the same area, and this so far as not been addressed by the current literature.

Finally, therefore the scope of this chapter has been to discuss not only the literature but, the policies and initiatives which drive the road safety campaigns in order to understand the current ways of thinking and how this is associated with the academic research literature. By trying to understand in greater depth the associations between road collisions and their risks, continued research would aim to understand this complex phenomenon.

CHAPTER 3

INFORMATION REQUIREMENTS FOR ROAD COLLISION AND SAFETY ANALYSIS

3.1 Introduction

Road traffic collisions take place at specific places and times. In other words they create 'snapshots' of time and space that can be used to evaluate the circumstances surrounding collisions. This representation of the real world is necessarily incomplete, and often inaccurate to unspecified degrees (see Longley *et al* 2005). As such, road traffic data are subjected to the same drawbacks as many other geographical data. These drawbacks fall under the umbrella term of uncertainty. While error and uncertainty exist in all spatial databases, it is the purpose of this chapter to determine whether and to what extent it has been managed in secondary road traffic sources. Primarily there are two types of uncertainty that are pertinent to this study the first of these can be applied to the spatial data. The spatial data is a secondary data source and often uncertainty in secondary data sources is unquantified. However it is important for the results of this thesis to evaluate and discuss issues of data quality and the sources and operation of uncertainties associated with these spatial datasets. This thesis primarily uses a UK road collision database called Stats19 which will be described in the next section of this chapter. In conjunction with this collision data a number of subsequent datasets have also been used including Mosaic scoring for driver and casualty postcodes (source: Experian, www.experian.co.uk), spatial locational data of speed cameras, schools, cycle lane location, pedestrian crossing data and road length.

This chapter aims to explore the nature of uncertainty associated with this analysis. Uncertainty as an umbrella term encompasses a wide range of concepts associated with road traffic collisions and the methods used in this analysis. This chapter is structured into two sections. The first of these addresses the uncertainty concepts surrounding the secondary Stats19 data source. It will incorporate an address of the Stats19 data and outline patterns and data preparedness. This will include general issues concerning the use and management of large spatial databases, as well as specific issues of how Stats19 is collected and modified with regard to data accuracy, data quality and under reporting. Discussion of the management of uncertainty in this section is limited to issues of awareness with respect to interpretation of results. The second section presents a more in-depth consideration of the most appropriate levels of scale for analysis and the prospects for sensitivity analysis using different geographical scales. Sources of uncertainty attached to methods used such as kernel density smoothing will be evaluated including the Modifiable Areal Unit Problem (MAUP) (Openshaw 1984) and network spatial autocorrelation (Goodchild 1986). Road collisions are placed in the genre of network analysis, where therefore they are constrained in space and time. This means that in order to analyse road collisions analyse of road collisions is usually with particular reference to the road network and its characteristics such as surface, speed limit and number of lanes. The road network reliability problems are rooted in the uncertainty of traffic conditions of the network. Many sources contribute to this uncertainty, ranging from irregular and random incidents, like earthquake, flood, adverse weather, traffic collisions, breakdowns, signal failures, road works etc, to regular fluctuations of travel demand in times of day, days of the week, and seasons of the year. Road collisions can only occur on the road network which makes the style of analysis particularly important with regard to whether they should be analysed on a network or with regard to administrative boundaries (which is how they are managed in terms of reduction initiatives) or other methods such as raster surfaces. This in turn will be evaluated in the chapter. In association with the analysis of these larger concepts, this second section will also approach the drawback associated with the temporal collision data and ways in which the merging of large datasets have been modified and managed. In summary therefore this chapter aims to discuss the types of uncertainty encountered within the substantive applications of this thesis. By addressing the different options available in terms of methods and how in turn each of these methods contributes to the overall uncertainty and then finally the choices which have been made and why.

SECTION A

3.2 Fundamentals of collision data

3.2.1 Description and discussion of the 'fitness for use' of the relevant datasets

The application of road collision and other relevant datasets (for example; traffic flow, demographic, land infrastructure and road use) are critical for the better understanding and management of road safety in general. Road safety analysis is often framed by the use of the road collision dataset, either a sampled specific dataset or police recorded collision data. This thesis uses the Stats19 road collision data which is recorded by the police at every reported collision. This introductory section of Chapter Three introduces the reader to the datasets used in this thesis, primarily the road collision dataset and subsequent, what I will call 'enrichment' datasets. What will follow, will be a preliminary investigation into patterns within the data and an evaluation of the strengths and weaknesses of the data, and if any modifications need to be made to it, and if so why. The first section will concentrate on the Stats19 database.

3.2.1.1 Definitions (severity, collision reporting etc)

Definitions should have been outlined in the initial report however; in order for the readers to have a sound understanding of the data I will outline some of the main definitions

Collisions that are not reported include:

Collisions which do not involve any personal injury
Collisions on private roads
Collisions reported to the police 30 days or more after it occurred
Collisions involving confirmed suicides

Severity:

Slight – An collision where the whole collision is deemed as having casualties who are suffering from sprains, contusions, slight cuts, shock.

Serious – Examples of serious injury are fracture, internal injury, severe cuts, burns, concussion, sever shock

Fatal – Where death occurs in less than 30 days as a result of the collision (does not include natural causes such as heart attack or suicide).

Within the London data certain variable are collected which are not officially required, these include:

Severity of collision (discontinued 1994)
Day of week (discontinued 1994)

Other variable information:

No of vehicle records – This is the number of vehicles that were involved in the collision

No of casualty records – This is the number of casualties injured in the collision

Collision reference number – For police use only in the event of a query

Local authority and police force codes – There is a list of coded numbers and which local authority and police force they relate to. With reference to the London data the police force code is useless because all the data has been recorded within the Metropolitan policing boundary. However the local authority codes can be deemed useful when outlining different boroughs and frequency of collisions.

Geographical coverage – specific reference to postcode data

Driver post code data – This is one recorded variable for each collision relating to the location of the driver's home address. The postcode varies between a complete six or seven digit postcode, a district or sector value and a blank value. The other point to note is not all postcode data will be able to be mapped because a percentage of the postcodes will be out of the London area of study.

Casualty post code data – Casualty postcode data is recorded within the table entitled 'Casualty details'. For each casualty there is an collision record, outlining details of post code (of which the same problems of the driver postcode apply). The other issue concerns the unknown frequency of foreign casualties that are not able to be recorded.

3.2.2 Preliminary Stats19 analysis

For each injury road collision known to have occurred in their areas, the police authorities complete a statistical return (which is called a "Stats 19" return), which provides details of the collision circumstances, separate information for each vehicle which was involved in the collision, and separate information for each person who was injured in the collision. Therefore the data is disaggregated into three tables according to the Stats19 records (see appendix for Stats19 data record sheets). Most of the variables are categorically coded and the codes are attached in the appendix of this thesis. These tables are segmented into: Attendant Circumstances, Casualty Details and Vehicle Details. These can be summarised below:

Attendant Circumstances: This section of the data records the general circumstances for the collision. There is one row in the dataset for each injury collision recorded. Data included in this table would include for example; geographical reference (Eastings and Northings), time, date, crash description, number of casualties, general level of severity etc.

Casualty Details: This dataset includes information specifically on the casualties of the collision, their age, gender, severity, whether they were a pedestrian, cyclist or car occupant etc. There is one row for every casualty recorded.

Vehicle Details: This table contains information regarding the type of vehicle(s) involved, information about the driver(s), age and gender and whether they were a casualty in the collision, speed limit.

A more detailed summary can be found in the Appendix.

There also consists a detailed crash description however there is no current universal language (for this description which means that it is problematical to do common word searches in the text to find common collision themes in this way.

The next part of this section shows results for various variables from the Stats19, within the context of the UK and London respectively.

3.2.2.1. Stats19 Age analysis

These three figures here (3.1-3.4) show a contextual summary of the age and gender disaggregation of the collision data. The age of the driver follows a coherent pattern, and in line with the idea of a large proportion of inexperienced drivers on the roads from the age of 17 upwards. The 17 and under age group will be largely associated with pedal cyclists or scooters/mopeds. There is a small peak in the age of casualties at 9-11 years old. This compared to London and the UK respectively stands out as a high risk age group compared to their population. The driver and casualty age data (Figures 3.3 and 3.4) also suggests rounding errors at the local peaks of 25, 30, 35, 40 etc. there may also be some potential data quality issues as it is surprising that as many 12 year olds as 75 year olds have collisions (see Figure 3.3).

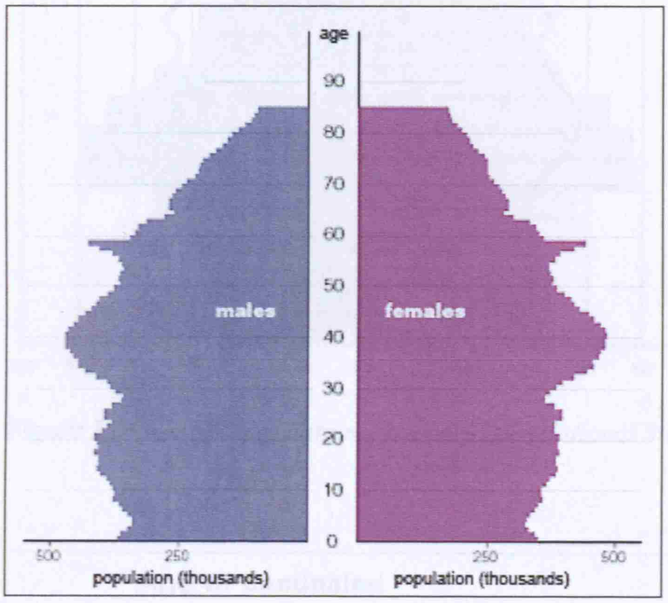


Figure 3.1: *UK population pyramid 2001 (National Statistics www.statistics.gov.uk 2005)*

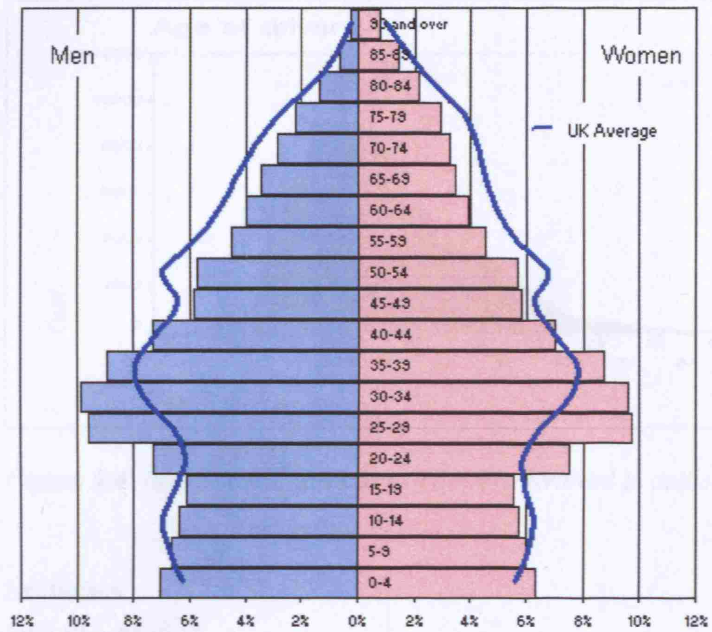


Figure 3.2: *London population pyramid 2001 (National Statistics www.statistics.gov.uk 2005)*

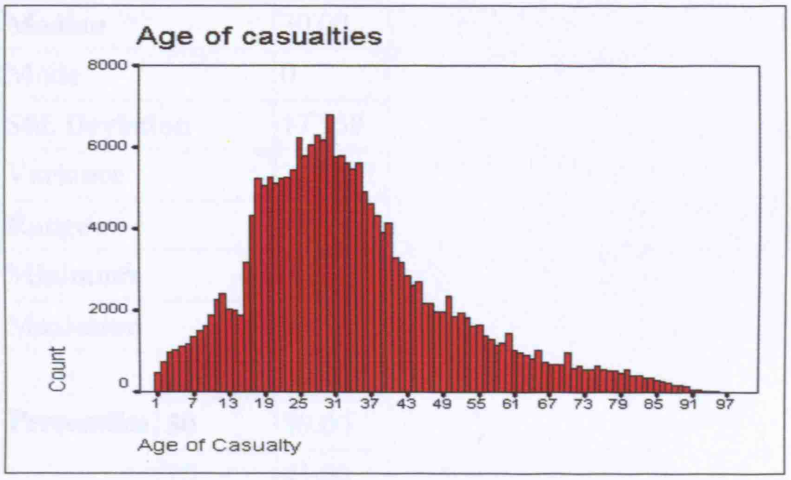


Figure 3.3: *Bar graph to show age of casualties involved in collisions*

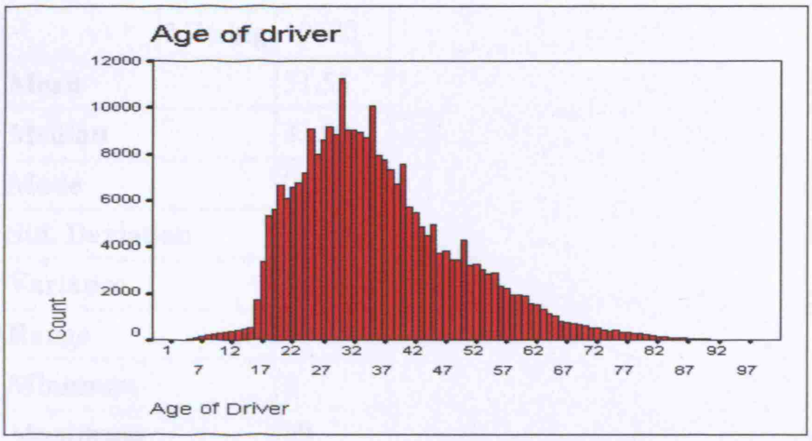


Figure 3.4: *Bar graph to show age of drivers involved in collisions*

Statistics Age of Casualty		
N	**Valid**	221789
	Missing	2693
Mean		32.12
Median		30.00
Mode		0
Std. Deviation		17.959
Variance		322.521
Range		99
Minimum		0
Maximum		99
Percentiles	**25**	21.00
	50	30.00
	75	41.00

Table 3.1: *Summary of statistics for age of casualty*

Statistics Age of Driver		
N	**Valid**	314436

	Missing	10620
Mean		31.53
Median		32.00
Mode		0
Std. Deviation		17.200
Variance		295.851
Range		99
Minimum		0
Maximum		99
	25	23.00
Percentiles	**50**	32.00
	75	42.00

Table 3.2: *Summary of statistics for age of driver*

These tables (Tables 3.1 and 3.2) illustrate information regarding the age of both driver and casualty. It is clear that from this information the mean age is similar for both variables and other statistics including the percentiles and range. These tables also report the number of missing variables in the data and one can clearly see that the frequency of missing variables for driver age is higher than for casualty age. However on closer inspection it is clear that there is a higher frequency for completed driver age fields. Another feature that these tables do not show that is clearly visible from the visual representations of the data is the high proportion of child casualties and the also elderly casualties.

3.2.2.2 Stats19 Gender analysis

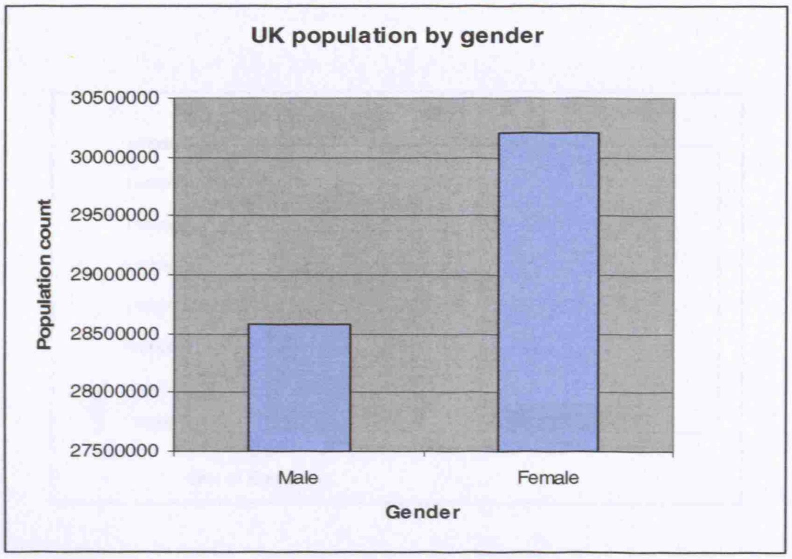


Figure 3.5: Bar chart to show UK population by gender

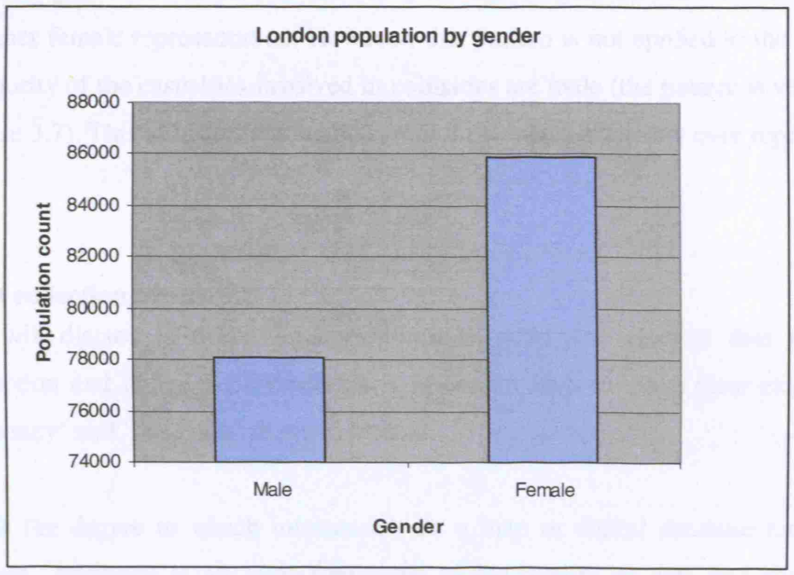


Figure 3.6: Bar chart to show London population by gender

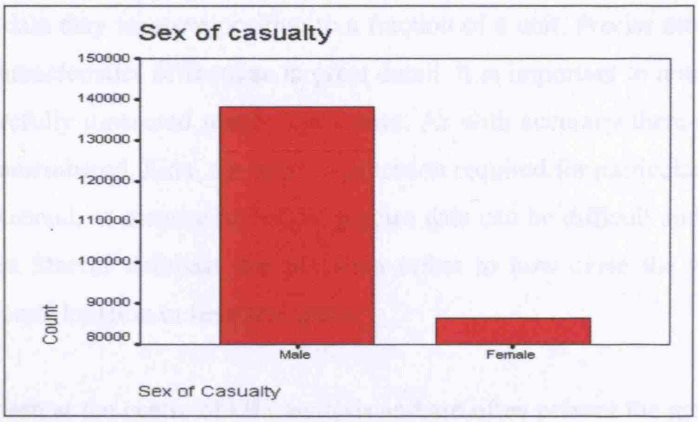


Figure 3.7: Bar chart to show collision casualty population by gender

These patterns of gender reveal both in the UK as a whole (Figure 3.5) and London (Figure 3.6) an overall higher female representation. However, this pattern is not applied to the collision data where the majority of the casualties involved in collisions are male (the pattern is very similar for drivers) (Figure 3.7). This indicates a deviation from the norm and a heavy over representation for males.

3.2.3 Current collection procedure

This section will discuss in detail the importance of good and accurate data for successful collision reduction and police deployment. It is important here to make clear exactly what we mean by 'accuracy' and 'precision' in terms of data:

Accuracy – Is the degree to which information on a map or digital database matches true or accepted values. Accuracy is an issue pertaining to the quality of data and the sources and operation of errors contained in a dataset or map. With regard to a GIS database it is possible to consider horizontal and vertical accuracy with respect to geographic position. It must be remembered that firstly the level of accuracy required for particular applications varies greatly. Secondly, highly accurate data can be very difficult and costly to produce and compile. In terms of the primary database used for the study, the accuracy of the Stats19 road collision database has not been officially quantified. It refers to the both the locational accuracy of the recorded collision location and its associated attributes.

Precision – Refers to the level of measurement and exactness of description in a GIS database. Precise locational data may measure position to a fraction of a unit. Precise attribute information may specify the characteristics of features in great detail. It is important to note that precise data no matter how carefully measured maybe inaccurate. As with accuracy there are certain points which should be remembered. First, the level of precision required for particular applications will vary greatly; and second, as mentioned before, precise data can be difficult and costly to collect. With regard to the Stats19 database the precision refers to how close the recorded collision location is to the actual location in time and space.

Data issues have been at the centre of GIS analysis and are often present the greatest impediment for any spatial application. If the data are not accurate then the GIS application results cannot be deemed to be accurate. Important decisions are made on the outcome of GIS applications and therefore they cannot be predicated upon inaccurate data. The importance of this theme cannot be stressed enough. Although it has been over 20 years since the wide advent of digital maps, global positioning systems (GPS) technology and machine searchable street names and co-ordinates, it can be challenging for some individuals to describe a location, and in the present context this problem is faced daily by the police when reporting collisions. Problems with location reporting have been observed anecdotally for a number of years, but there has been relatively little scientific documentation on the scale of the problem and its solution.

At this point it is important to stress that individual police authorities have autonomy in deciding what if any data they collect beyond the Home Office Requirements. Police forces have plenty of opportunity to collect data pertaining to their own forces, yet the data gathering exercise is not systemised between police authorities and therefore data are often not geographically comparable between authorities. The important point to make here is that the causes and consequences of road traffic collisions do not respect the artificial boundaries of police jurisdictions.

3.2.4 The need for good data

GIS in a policing and engineering environment is used as a decision support tool (Veregin 1989). The role of GIS as a spatial decision support tool has been widely acknowledged but its potential in the role of collision reduction and traffic policing is only slowly being realised in road policing applications. In any data collection process there is a large margin for error and it is the ways in which such error is reduced through various filters and checks that is important. Michael

Goodchild (see Zhang *et al* 2002), a leading academic in the field of spatial uncertainty and inaccuracy comments:

- All spatial data are limited in accuracy
- Available precision in GIS software systems exceeds the accuracy of data
- The means to characterise the accuracy of spatial data, track uncertainty through GIS processes and compute and report uncertainty are inadequate.

3.2.5 Sources of data uncertainty in measuring and analysing road traffic collisions

Following Guptill (1989), in order to achieve better data and more reliable decisions, the following points should be remembered when discussing the importance of data accuracy and quality within the police

- Improvement of models of uncertainty
- Methods of encoding uncertainty in databases
- Methods of tracking uncertainty
- Methods of computing and communicating error in products and policies

Recently the term 'data quality' as been replaced by uncertainty (Buttenfield *et al* 2001: although this does not mean we should discontinue using the term 'data quality'). 'Uncertain' does not necessarily mean unreliable, changeable or erratic (Oxford 1996 cited in Buttenfield *et al* 2001). Uncertain data possess attributes of either accuracy (an affirmative attribute, measured in terms of similarity) or error (a negative attribute, measured in terms of discrepancy).

Recording of road traffic collisions is a process which is highly vulnerable to uncertainty and is important because the objective would be to minimise the time and cost spent on policies which derive from inaccurate and uncertain data. Listed below are some of the various types of uncertainty which may occur within road traffic collision data collection and analysis:

- **Measurement error** - whereby the location of the collision is recorded either as textual information or later encoded into an eight figure grid reference or is recorded incorrectly using a grid reference at the scene of the collision. Most police officers attending a collision do not have GPS equipment to make accurate recordings of location.

- **Uncertain boundaries** - Because there are few 'natural' units of road traffic collision measurement this provides uncertainty in terms of boundaries. The boundaries are created on an individual force basis and, in the absence of universal guidelines these are usually based on policing units. However this proves inaccurate because collisions often do not conform to boundaries because their causes and consequences have a much wider scope (this will be discussed in more depth in Section 2)

- **Data transformation** - From the example of the Metropolitan Police collection process (Section 3.3.1), it is apparent that there is uncertainty at the stage where the textual data are transformed into a grid reference since this transformation relies on a number of factors in order to be precise. These include the accuracy of the textual information and how well the encoder can interpret it to a grid referenced map. Such uncertainty arises principally because of data quality issues within an organisation that can mean the difference between meaningful maps and a simple graphic or poster. One must be aware of the fact that high data quality is not necessarily data that is devoid of errors. Incorrect data is only part of the data quality equation and even incorrect data can go undetected in data quality filters. In the realm of road traffic collisions the organisation that is collecting the data is the police, and they are not necessarily the only other organisation to be using the finished data set so data quality is even more important because of the number of partners (for example Transport for London, all London borough councils, independent analysts, universities, health professionals and road engineers) involved in reducing collisions.

- **Age of data** – The last collision to be recorded was 30th March 2003. The first incident recorded was January 1998.

- **Geographical coverage** – The data covers the police forces and local authorities in London (they are geographically coordinated). Therefore it is apparent that there may some issues surrounding incidents that have occurred on border roads between authorities both between London and outside London. Not all of the Eastings and Northings lie within the study area. The map in Figure 3.8 shows the overlay of London's borough boundaries and the collision points.

then interpreted by a team of people into a grid reference. There is a high error margin for this process.

- **Sources of variation in data** – It is clear from discussions with Road Traffic Police Officers that data is collected by the attending officer and then inputting into the computer database by a number of people; however because of the Stats19 recording practices any error should be minimized.

- **Classification problems** – There is a classification problem when concerning time and date however this will be discussed in the next section. In terms of post code data there is discrepancy of incomplete post code data and an unknown postcode where a one digit number is given. The code was at the discretion of each police authority and therefore is meaningless.

3.3 The Stats19 collection system

Personal injury data arising from road collisions were first collected in 1909. The modern 'Stats19' collection system was established in 1949 and the current collection system was implemented in 1979 after a wider ranging review. Road collision statistics are essential for informing and monitoring road safety policy at local, regional and national levels. Locally they are used to support remedial engineering work on public roads. At a local and national level they are used to underpin road safety strategy and targeting casualty reduction. Individual police forces and local authorities require road collision statistics to support their own road safety policy programmes, which vary in focus (from child pedestrian safety to tackling drink drivers). The collection process and data collected vary between local authority and police force areas, reflecting a range of different road safety requirements and initiatives. London is unique because all the road collision data is collated by a central body (The London Road Safety Unit) and hotspots are then sent out to each borough for further analysis and remedial measures. However each local area is required to report the same set of collision records for national purposes and to transmit them to central government. These are what are known as Stats19 records.

The accuracy and credibility of the Stats19 collection process depends upon close co-operation between central government, local government and police forces. The Stats19 system is jointly

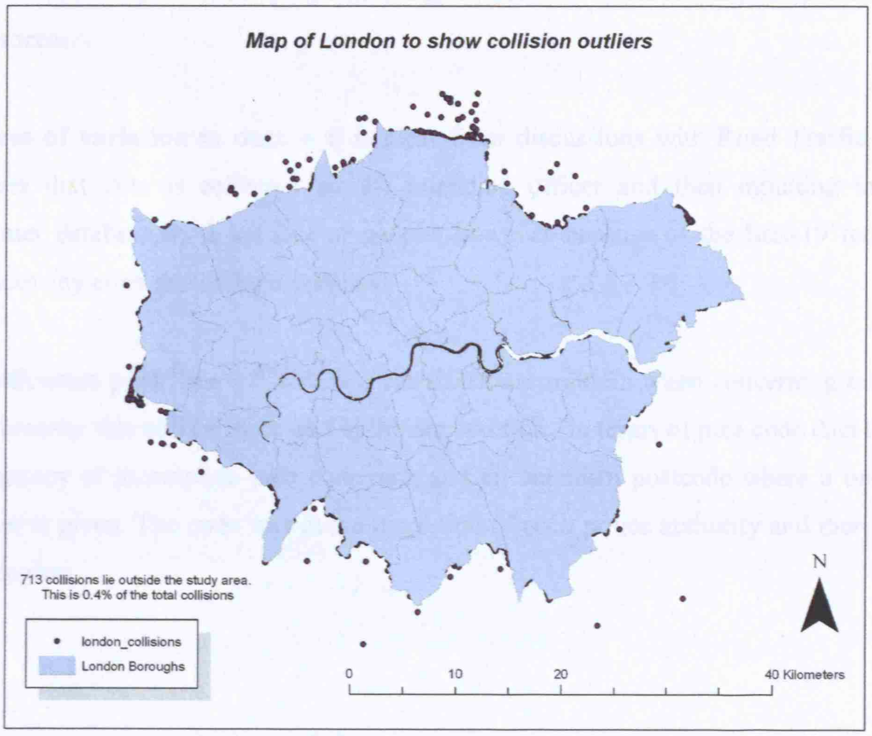


Figure 3.8: *Map of London to show collision outliers*

- **Data scale** – The data is referenced as an eight figure northing reference and an eight figure easting reference. Therefore the incidents are measured to the nearest 10 meters which is the norm for any collision data collected in the UK in conjunction with the STATS19

- **Relevance** – The data collected consists of an increased number of engineering and road variables due to an increase in road engineering safety measures). There is little information regarding the drivers and casualties and the area in which the collision took place, such as 'built up, industrial, residential etc' which would be helpful. As far as driver and casualty data is concerned, this dataset has post code data (some only textual post codes) which is helpful as it is not a requirement within the STATS19 requirement form.

- **Positional accuracy** – Collisions are recorded in terms of a textual description by a police officer at the scene of an incident or description by a second hand source. This is then interpreted by a team of people into a grid reference. There is a high error margin for this process.

- **Sources of variation in data** – It is clear from discussions with Road Traffic Police Officers that data is collected by the attending officer and then inputting into the computer database by a number of people; however because of the Stats19 recording practices any error should be minimized.

- **Classification problems** – There is a classification problem when concerning time and date however this will be discussed in the next section. In terms of post code data there is discrepancy of incomplete post code data and an unknown postcode where a one digit number is given. The code was at the discretion of each police authority and therefore is meaningless.

3.3 The Stats19 collection system

Personal injury data arising from road collisions were first collected in 1909. The modern 'Stats19' collection system was established in 1949 and the current collection system was implemented in 1979 after a wider ranging review. Road collision statistics are essential for informing and monitoring road safety policy at local, regional and national levels. Locally they are used to support remedial engineering work on public roads. At a local and national level they are used to underpin road safety strategy and targeting casualty reduction. Individual police forces and local authorities require road collision statistics to support their own road safety policy programmes, which vary in focus (from child pedestrian safety to tackling drink drivers). The collection process and data collected vary between local authority and police force areas, reflecting a range of different road safety requirements and initiatives. London is unique because all the road collision data is collated by a central body (The London Road Safety Unit) and hotspots are then sent out to each borough for further analysis and remedial measures. However each local area is required to report the same set of collision records for national purposes and to transmit them to central government. These are what are known as Stats19 records.

The accuracy and credibility of the Stats19 collection process depends upon close co-operation between central government, local government and police forces. The Stats19 system is jointly owned and managed by the Standing Committee on Road Collision Statistics (SCRAS). In England, within each local area, Stats19 data are collected by a central unit referred to as the Local Processing Authority which can be managed directly either by a police force or a local authority, or be subcontracted to a private consultancy. There are 58 Local Processing Authorities in Great Britain of which just under half are managed by local police authorities and the rest by local authorities. The Stats19 report form consists of a collision record, a vehicle record to be completed for each involved vehicle and a casualty record for each casualty arising from the injury collision. In 2000, local authorities and police forces collected, coded, validated and reported 234, 000 collision records, 430, 000 vehicle records and 320, 000 casualty records for central government (see appendix for a copy of the Stats19 record sheet).

3.3.1 Example of road collision data collection: The Metropolitan Police

The map below (Figure 3.9) shows the different boroughs that combined together form the Metropolitan Police Force Area.

Figure 3.9: *Map of police force divisions within the Metropolitan Police*
Source: http://www.met.police.uk/about/boromap.htm (20.05.03)

Most police forces (of which there are 43 in England and Wales) have a different verification procedure of the collision data. Within the London Metropolitan Police as with all other forces, they only record collision locations to which they are called. This means that many collisions go unrecorded but this will be discussed in the next section. The following steps are taken to record a collision and its location.

Police officer(s) will attend a collision and will record responses to the questions from the Stats19 completion form in a small black logbook. There are eight boxes for an eight-figure grid reference however it would be naïve to assume that police officers would know the exact location of an collision so the attending officer gives an interpretation of the location (usually in the form of landmarks and road markings).

- The paperwork is then photocopied three times and two copies are filed and stored within in the police borough and the other is sent to the London Processing Unit so it can be entered on a database and a grid reference can be determined from the notes. This clearly has a high margin for error and mistakes could occur. It will rely on interpretation from the written transcript to relate it to an Ordnance Survey map.

- The database is then sent to the London Collision Analysis Unit for analysis at a basic level and a breakdown at a borough by borough level. This is one example of a police force and is unique because it is London's collision reporting system and is responsible for an area of high population density and a large number of collisions. Other forces have different agendas and reporting units.

3.4 Under-reporting and accuracy of collision data

Not all road collisions are recorded by the police. The collisions recorded by the police are likely to be injury collisions however even some of the more minor injury collisions do not get reported. This is called under-reporting. The under-reporting of traffic collisions is considerable and depends on the type of collision. It can also vary from one force to another. It is a problem when it comes to having a complete picture of the magnitude of the traffic accident problem. The most serious accidents are nevertheless reported almost fully to the police. Drivers involved in single vehicle collisions are less likely to report collisions. Elvik *et al* (1999) compared 49 studies in 13 countries and showed that the reporting of injuries in official accident statistics was incomplete at all levels of injury severity. The mean reporting level of collisions in all countries was found to be 95 percent for fatalities, 70 percent for serious injuries requiring hospitalization, 25 percent for injuries in which the patient was treated as an outpatient and 10 percent for very slight injuries.

Before the nature of under-reporting and accuracy is discussed it is important to differentiate between the two terms as a certain amount of blurring in the use of the two terms has occurred in recent years. A primary source of the under-reporting of collisions concerns the legal requirement that only collisions in which a motor vehicle is involved causing injury to a person other than the driver must be reported to the police (James 1991). Therefore there are a large number of collisions that go unreported for the many reasons which will be discussed later on in the section. Secondly accuracy concerns how correct the information is regarding the road traffic collision.

James (1991) presents a discussion of the under-reporting of road traffic collisions and concludes that collisions involving children are more likely not to be reported to the police. Ibrahim and Silcock (1992) focus on the suggestion that the accuracy of road traffic collision data is reduced because of the high levels of under-reporting that occur, and suggest that the collision data are as a consequence misleading because they do not portray all of the collisions (this is especially true because police only collect data on road traffic collisions that are reported to them) Thus their data do not include minor collisions that may only result in an insurance claim. There has been little or no attempt to estimate the number of un-reported collisions that occur, and the only practicable way to do so would be to investigate insurance companies and hospital in-patient records. Ibrahim and Silcock (1992) discuss how much time is spent checking the collision data (although primarily this is in the hands of the individual police forces). The article states that:

'The inaccuracy of collision location by the grid reference and the plain language description are the two most frequent problems'

(Ibrahim and Silcock 1992)

It seems clear that even ten years ago, the accuracy of road collision data was an issue; however the follow up has been unsatisfactory. James (1991) attempted a broader study of data which entailed comparison of hospital statistics and reported road traffic collision statistics. It is also possible to outline possible uncertainties which surround this area of research, for example because road traffic collisions have no 'natural' units of analysis and studies of the spatial patterns of road collisions all use different spatial units, making it difficult if not impossible to establish relationships between areas. James (1991) remarks:

'In addition to under-reporting, the problem of misclassification of information by the police was also apparent. Assessment of injury severity was often inaccurate.....' (James 1991, Traffic Engineering and Control)

From the evidence it is clear that the theme of accuracy and under-reporting has been acknowledged for a number of years, yet there has been little work which has built on these foundations. There has been minimal data sharing between insurance companies, who record a large proportion of the damage only collisions some which are not reported to the police. A proportion of police forces do collect information of damage only collisions and they are recorded in a separate database. However they are rarely used for collision investigation as no grid references are assigned, hence the accuracy of the locational variables relating to these records cannot be validated (Austin 1995). This information from insurance companies is researched privately and the possibilities of data sharing with the police would assist in the reduction of collisions and collision hotspots.

3.5 The police and geo-referencing

The restricted nature of police budgets and enforcement of government policies, the collection procedure of road collision data means that information can be limiting and occasionally dated. As mentioned in the example of the metropolitan police, the reporting system has very serious implications for error margins. It is inevitable mistakes will be made because the police have many duties to perform at the scene of a collision (Austin 1995). All collisions have the potential to have an eight figure grid reference, and all the data processed do possess an eight figure grid reference. The grid reference is obtained after the event based on a locational description by the attending police officer(s). This textual description and its interpretation could potentially lead to an error in the location of the collision. The problem has been acknowledged by the Association of Chief Police Officers (ACPO) and other policing organisations; however it would be inaccurate to assume police officers could work out an accurate grid reference for each collision they attend without having the use of an Ordnance Survey map or Global Positioning System (GPS). One of the possible ways this problem could be solved would be with the use of Global Positioning Systems (GPS) or Ordnance survey maps.

3.5.1 The potential of GPS

The use of GPS within GIS is a very powerful tool as it enables users to gain spatial information about most parts of the world, at almost any level of detail (Kennedy 2002). If a differential GPS is used (where two receivers are used) the accuracy can be up to 10cm, and even a stand alone GPS will have accuracy of approximately 5m. GPS can serve as a means of data input for GIS. GPS provides users with a convenient method for assigning and using absolute co-ordinates. Therefore people can now know their positions and combined with a map they can then know their location (i.e., where they are in relation to other objects around them). Only a few police cars are equipped with GPS receivers. The accuracy of GPS is very high and the collision would be measured at a 5 metre resolution. In a recent research report by the Kentucky Transportation Centre, an analysis is made between using GPS and their own mile point data collection procedure. The results produced highlighted that the GPS located the collisions with more accuracy and the only margin for error was operational rather than equipment or environment.

There are a few locations that GPS cannot be used, however the only one that would be applicable to road collision data would be that of underground (so collisions in tunnels would be restricted to police officer recording), and in 'urban canopies' amongst tall buildings. In practice this would probably account for large swathes of urban areas where collisions are disproportionately concentrated. Without going into too much technological detail regarding GPS, it is more important to understand the potential for GPS within the realm of road traffic collision data and how accuracy could be increased and error decreased by implementing such a tool.

3.6 Categorical data error

The majority of data entered into the STATS19 database about the circumstances of the road collisions are in categorical data format. These data are entered into the computer after the event and are therefore subject to the inconsistency of recording. This human observation is heightened by the trauma of attending a road collision event and noting down the specific details regarding for example the weather – 46% of the collisions in the STATS19 'Attendant Circumstances' dataset were categorised as having weather 'unknown'. Thus in nearly half of this dataset the weather at the scene of the collision cannot be confirmed, highlighting a large margin for error in trying to predict the type of weather in which more collisions occur in than others. Only one value can be entered for each item which can be restricting if two features are present at the collision site for example traffic lights and a roundabout.

Coupled with the inconsistency in human observation is the complexity of the distribution of categories. The categories in the STATS19 database have evolved over many years and are constantly being updated and changed in order to enhance the quality of information collected. The 2003 Quality Review addressed some major changes such as the deletion of some variables and the addition of others. It also has tackled challenges such as the availability and linkages with other socio economic data. The most significant change occurred to the addition of contributory factors which meant that the causal factors associated with the collision would be documented, including information on the purpose of the journey, the mood of the driver and the effects of distractions. All of these are difficult to quantify and outline using the existing variables. Zhang *et al* (2002) expressed that categorical data should be conceptualised as being spatially varying. This means when spatial data is recorded (such as road collisions) by different people at different places their perception and use of the categories may differ creating a spatial uncertainty.

Managing categorical data error from data sourced from third parties can be difficult. The STATS19 database is vigorously validated and this validation system can identify nearly all of the mistakes (Austin 1995). However those collisions that are coded incorrectly will alter the number of collisions relating to a certain feature whilst those that are wrongly located will alter the number of collisions at certain sites and therefore reduce the validity of hotspot investigation analysis.

3.7 Temporal and diurnal patterns and sources of error

This section introduces the reader to the concept of temporal patterns within the Stats19 data, by investigating the data in an exploratory level is was evident that reporting error was present and a need to overcome this.

Date – The date format is {01-Jan-98} and ascending to {31-Mar-03}. A frequency table has been established to analyse the rates of collisions over the 63 months worth of data.

Time – The time is has been recorded to the minute in the format {30-Dec-1899 18:30:00}. However after looking at the frequency of the data it is clear that the majority of the collision times have been recorded to the nearest five minute interval because there are an increased number of collisions occurring at every five minutes. Therefore, initially to overcome this problem it would be firstly beneficial to group the times into hours and then as analysis continues break down this down into possible ten minute averages, because otherwise the time analysis would be obscured by the database entry error.

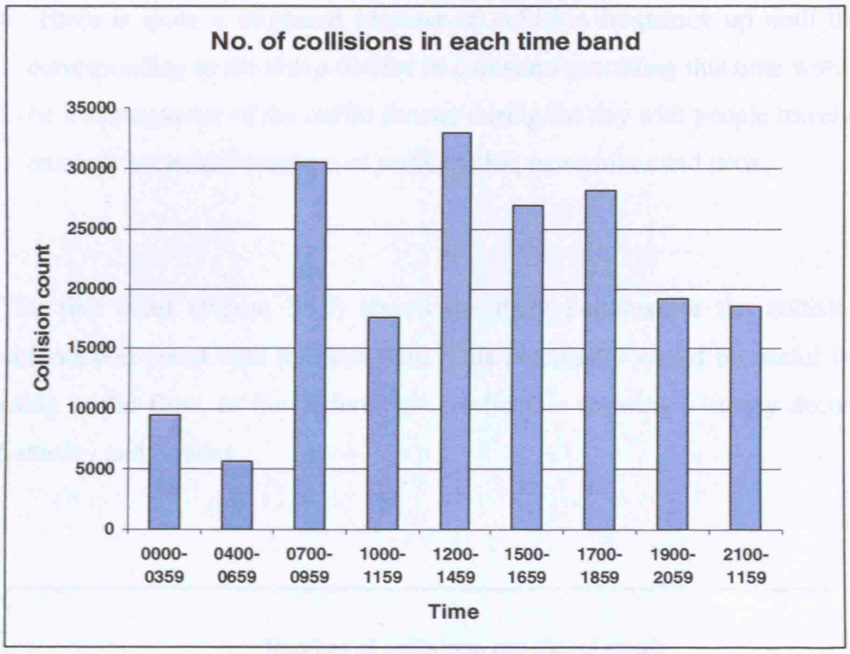


Figure 3.10: *Number of collisions in each time band*

The illustration above (Figure 3.10) shows the frequencies of collisions broken down into temporal groupings starting from 1200-0100 and onwards. The following points from this frequency graph are listed below:

- Figure 3.10 shows the general pattern for collisions over the course of a day. There are two clear peaks, the first of which is between 0700 and 0959, characterised by the heavy traffic and commuting. Unfortunately, it is not possible to create a rate of this data against traffic flow which would possibly give a more accurate indication of the diurnal patterns. The second noticeable pattern, and the peak is between 1200-1459, which is interesting as it would be predicted that the two commuting periods (0700-0959 and 1700-1859) would share the two highest counts of collisions.

- There is a considerable dip in collisions occurring between 0100 and 0500. This will be due to the reduced traffic density on the roads, however further analysis would suggest that these collisions could possibly be more severe due to lower traffic density and vehicles driving at higher speeds.

- There is quite a staggered increase in collision frequency up until the peak at 1700-1800, corresponding to the sharp decline in collisions preceding this time within London. This would be a consequence of the traffic density during the day with people travelling around the city for example increased numbers of pedal cycles, motorbikes and taxis.

The first chart (Figure 3.10) shows the daily patterns for the collision data. It shows the unformatted count data for each data. This potentially would be useful to transpose into a rate using traffic flow, as the patterns are predictable showing a steady decrease in collisions on a Saturday and Sunday.

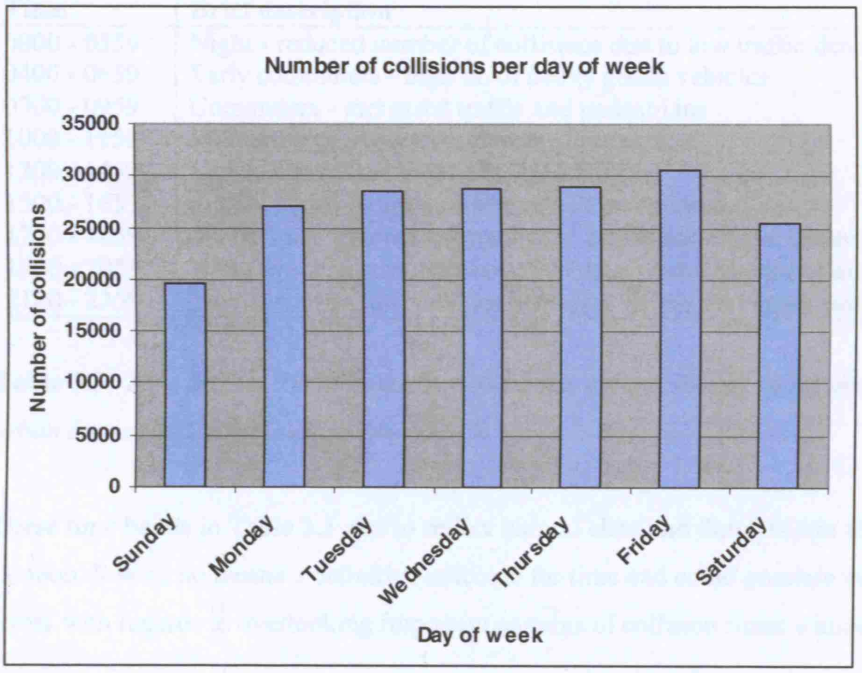


Figure 3.11: *Number of collisions per day of week*

Temporal accuracy for road collision analysis is very important. By knowing the time of the collision it is possible to relate it to other environmental and traffic databases. However the Stats19 database used for this study highlighted some temporal inconsistencies. There are two margins for error in the temporal variable, the first of which cannot be managed but merely speculated upon. For every collision a time is recorded indicating the time at which the collision occurred. This information is recorded by collating information from people involved in the collision, from the officer at the scene and from the first emergency telephone call. In retrospect, this may not be accurate, and there is likely to be some error in the time of recorded collision because it is likely there is to be a time lag between the time the actual collision occurred, the time the police are called and the time the police arrive.

The second margin for error was identified when the times of the collisions were looked at more closely. Every time is recorded on a twenty four hour time and seemingly to the minute. However when the times were analysed more closely, it showed significant peaks around every five minutes. This shows that the officers recording the collision time were inclined to round up or down to the nearest five minutes whereas other police officers would try and note the exact time. In order to manage the error, time bands were created (Table 3.3)

Time	Brief description	Code
0000 - 0359	Night - reduced number of collisions due to low traffic density	1
0400 - 0659	Early commuters - high no of heavy goods vehicles	2
0700 - 0959	Commuters - increased traffic and pedestrians	3
1000 - 1159	Mid morning - couriers, increased pedestrians	4
1200 - 1459	Lunch - Increased number of pedestrians	5
1500 - 1659	School finish - high numbers of school children	6
1700 - 1859	Work finish - increased number of traffic and erratic behaviour	7
1900 - 2059	Early evening - less number of vehicles, reduced pedestrians	8
2100 - 2359	Late evening - less vehicles but faster speeds and drunk pedestrians	9

Table 3.3: *Time bands: These bands represent the natural breaks in the day associated with an urban auto centric area such as London.*

These time bands in Table 3.3 aim to reflect natural ebbs and flows within the daily flucations in London. It is by no means a definitive outcome for time and could possibly introduce other bigger errors with regards to overlooking important patterns of collision times within the time bands.

3.8 Linking road collision databases to other transportation and traffic data

This study requires ancillary datasets in order to carry out a more in depth study into the causes of road collisions and creating a meaningful typology. A strategic aim for this study was to create a database for all identified hotspots, using information about the road environment which is not available in the Stats19 database. This included information on cycle lanes, speed cameras, schools, bus stops, pedestrian crossings and tube locations. Different units and agencies collect traffic volume, roadway inventory, and land-use data. For example, road network data come from the Ordnance Survey, cycle lane data come from London's Cycle Network, and bus stops, pedestrian crossings and traffic lights come from the Department of Transport under the Freedom of Information Act 2005. The multiple data sources pose numerous consistency and currency problems. Assembling all this information for the planning area may be difficult but would add to the quality and depth of collision analysis.

The ancillary data was linked to the hotspot sites using GIS (specifically ArcGIS). Using the hotspot basic spatial unit of the 100m² cell (refer to Chapter Five) it was possible to total the length of road network and cycle lane network in each hotspot. This was achieved by using the 'spatial join' function in ArcGIS. For the point data (schools, bus stops, traffic lights and tube location) the same spatial join procedure was used to find the total in each hotspot. The pedestrian crossing data was in the format of a polyline shape file which meant a spatial join could not be achieved between two polyline files therefore this procedure was done manually for each hotspot.

Linking data reported in different ways creates another problem. Traffic collision data are identified by specific locations, usually represented by points on GIS. Traffic volume data are usually measured over specific links of a highway network. Using GIS, these data can be represented by line segments or, occasionally, by points (to measure, for example, the mid-point of a segment). Road inventory data (types of roads, lanes, bridges, rail lines) can be represented as lines or points. On the other hand, traffic analysis zones (TAZs), the basic analysis unit of travel-demand forecasting, are represented by zones or polygons on a map. Land-use data are frequently represented by zones but can also be represented by lines (block faces) or points (specific buildings). All data types that may be included in a collision information system can be measured in different ways and represented in different geometrical units. Added to this is the complexity involved in translating these data sources into the same geographic coordinate system (for example, state plane coordinates). Linking all of these data requires a complicated set of programs and routines. In addition, other tools and programs must be linked to this data so that analysis can be performed.

SECTION B

3.9 Collision data analysis, tools, techniques and sources of uncertainty

3.9.1 Introduction

This section moves beyond discussing the collision data and its limitations to discuss the sources of uncertainty and scale in the actual analysis of the data itself. This section is disaggregated into three parts; this first of which highlights the nature of the basic spatial unit (BSU) for analysing road collisions and how this is approached in different studies and the justification for the choice of BSU in this study and how it is created. It is a continuously contentious issue due to a road collisions spatial constraint on the road network and the lack of consensus over which BSU to use. This section also addresses the limitations of using different BSU's and how this relates to the chosen method to create the BSU, kernel density estimation. The second section closely linked to the first discusses the importance of scale. Multi level models have become important analysis tools in road collision investigation in recent years (see Jones *et al* 2003, Eckhardt *et al* 2004). In particular how the characteristics of space (environment and infrastructure) can influence the location of collisions at different levels of measurement (Erkhardt *et al* 2004). Therefore it highlights the importance of the size of the BSU rather than the type outlined in the previous section as a source of contention for the most accurate analysis of road collisions for this particular study of London. The final section relates to the ranking of hazardous locations or hotspots which traditionally have been based on severity (see Geurts *et al* 2003, Brijs *et al* 2004). This section explores other methods of ranking these sites based of different criteria and the choices and justification for the ranking method in this study.

3.10 Spatial/network autocorrelation and MAUP

The fundamental objective of spatial analysis of road collisions is to reduce them. The methods and procedures of spatial analysis can take many different guises, most obviously based upon mapping using GIS. A road collision in this context is often seen as a point or occurrence on a map. However in many modelling and prediction techniques traffic collisions tend to be aggregated either to links on the road network or to administrative units. This then leads to a fundamental problem experienced by many geographers, which is determining, or working with, the most appropriate size and shape of the spatial units used for analysis, since this may heavily influence the visual message of mapping and the outcome of statistical tests. In this context, Thomas 1996) explore aspects of Modifiable Areal Unit Problem (MAUP) with the objective of finding the optimum length of road segment for road collision analysis in Belgium. Both Yule and Kendall (1950) and Openshaw and Taylor (1979) recognised the size and scale problem, which is present in many geographical studies with spatial datasets. Generally, when increasing the level of granularity of the analysis, for example from London boroughs to Census output areas for London (in terms of road collision count data summaries) the correlation between Output areas will become weaker. This introduces another issue surrounding road collision analysis and that is spatial autocorrelation and in recent years network spatial autocorrelation. Spatial autocorrelation refers to the extent to which the value of a variable at a given location influences values of that variable at contiguous locations (see Cliff and Ord 1973, Griffith 1987, Goodchild 1986 and Odland 1988). For example this could be the influence of a variable on road segments that lie next to or close to each other or grid cells with count or frequency road collision data and their associated likeness. The notion of spatial autocorrelation rests on the premise of Tobler's 'First law of geography' (1970) which determines that 'everything is related to everything else but near things are related more than distant things'. Spatial autocorrelation simultaneously deals with both the attributes of the spatial data and the spatial feature as a location; in other words, it deals with the location of the collision itself and the attributes of the collision such as time or severity.

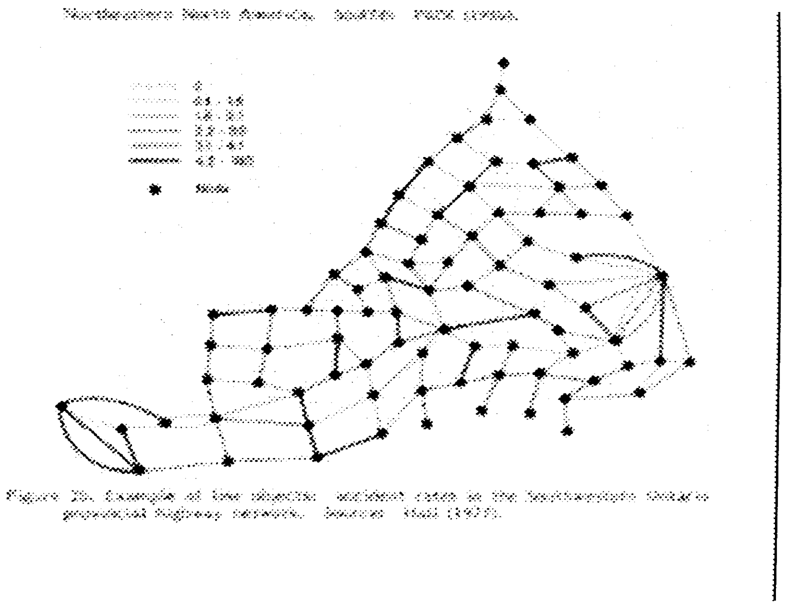


Figure 3.12: *Network spatial autocorrelation (source: Goodchild 1986)*

Figure 3.11 shows collision statistics on road links on the Ontario provincial highway network. Low spatial autocorrelation in these statistics would imply local causal factors such as 'hotspots' whereas strong positive autocorrelation would imply a more regional scale of variation pointing to causal factors such as lifestyles, and rural and urban land uses.

One of the fundamental aims of this thesis has been to identify road collision hotspots. In order to identify the most suitable method there are number of considerations to be taken into account such as the scale at which a hotspot is deemed manifest, the basic spatial unit (for example a section of road) and the boundary problems associated with any determination of a basic spatial unit. Flahaut *et al* (2003) argue that the spatial aspects of road collisions have often been neglected in the literature and that the several basic methodological aspects such as the definition of a hotspot are still held in low esteem particularly the spatial aggregation of the data and the definition of the concentrations of road collisions. It is evident from the literature however (see Thomas 1996, Flahaut *et al* 2003) that because there are spatial concentrations of road collisions there is a spatial dependence and interaction occurring between contiguous collision locations.

The problem arises that when hotspots are determined, whether they consist of one basic spatial unit or twenty, is that it is often assumed that these high concentrations of collisions are spatially independent of each other (albeit that the collisions in the hotspot are spatially dependent because they all occur in a similar spatial area). In other words the cause of the hotspot is mutually independent compared to other areas of high collision concentrations. Therefore it has been necessary in this thesis to not just ascertain the dependence of each collision within the hotspots but to identify similarities of the hotspots themselves across London. In this analysis kernel density estimation has been chosen in order to determine the spatial density of the collisions. The discussion of the kernel density parameters will be discussed in chapter five however it is important at this point to mention that the choice of parameters will affect the location of the hotspot. The analysis of road collision hotspots can be aggregated into different BSU's for example a length of road network and a grid cell (generated from kernel density estimation). This is known as an aggregation problem and is part of the modifiable areal unit problem (the other being scale which will be discussed in the next section). The aggregation refers to the shape of BSU of the hotspot in this case. There are two main choices which will be presented here, the first being road network segments the other being grid cells. Both share aggregation drawbacks. The road network neglects information regarding to schools, shopping centres, and land use. In comparison grid cells can include socio economic data not recorded on the road network as well as the total length of road network occurring within the grid cell(s). Many of these socio economic variables are not related to the road network environment. Therefore if road network segments were to be used it would not encompass the surrounding infrastructure and environment associated with the possible causes of a road collision hotspot.

In this study there are two scales of spatial autocorrelation working together. The first is the spatial dependence between the collisions within the hotspot, and the second is the clustering of different hotspots. Research by Thomas (1996) has highlighted the utility of network spatial autocorrelation measures as a way of determining the location of high collision locations as road segments. Thomas (1996) defines positive network spatial autocorrelation as being a discernable clustering of network segments having similarly high collision rates. A discussion of the limitations of network segmentation and why it was not specifically used in this thesis is discussed in depth at the beginning of Chapter Five. This application can be adapted to the current study, by taking grid cells which experience a similar high density score with respect to collisions. Spatial autocorrelation of collisions amongst cells that are close to one another leads to their grouping and classification as a hotspot. The number of grid cells that may be grouped together can vary and therefore leads to scope for enabling cross classifications with other hotspots that share similar characteristics including size and shape. This means the method is not subject to a uniformity of shape and the hotspot can be a very large number of cells or just one cell.

This study will use kernel density estimation rather than the standard frequency count data that many London boroughs use, a collision frequency over three years. The kernel density estimation method means that the spread of a collision is over a larger area not just the point itself. This implies a spread of risk around any given cluster of collisions. In a spatial sense this may seem more in - accurate and it is not devoid of uncertainty. Due to the uncertainty of the boundaries created from this method it may be the case that they are not the most accurate boundaries around the cluster of collisions. This level of inaccuracy is present in many different collision clustering methods such as Thomas (1996) network segmentation method and it is often difficult to determine every collision in a spatial area that may share some common characteristic. In short, the spatial clusters may tend to share a common cause explaining the reason for the hotspot. Surely it may be the case that it is only spatial characteristics they share and nothing else? The aim of Chapter Six is to ascertain the level of attribute dependence which is associated with the collisions in the hotspot.

3.11 Scale

Measurements of spatial and temporal autocorrelation are scale dependent (Longley *et al* 2005). This scale dependency has limitation when applied to spatial analysis. Scale is also part of the modifiable unit problem (MAUP) whereby different results can be obtained by changing the spatial resolution of the statistical analysis. The Modifiable Areal Unit Problem (MAUP) is a potential source of error that can affect spatial studies which utilise aggregate data sources (Unwin, 1996). The effects of the MAUP can be divided into two components: the scale effect and the zonation effect (Armhein 1995). The use of the basic spatial unit for the hotspot analysis is affected by both components. For the purposes of this thesis the meaning of scale will be taken to mean the spatial resolution of the analysis (particularly referring to the basic spatial unit of the cluster analysis) and the extent of the analysis in geographical terms. The basic spatial unit which will be used for the hotspot analysis is a 100m^2 cell. The rationale and reasoning for this will be discussed in depth in Chapter Five. The modifiable areal unit problem in this context (scale) means that if the cell size is increased the results will be altered.

The possible existence of hierarchical structures within road collision data is commonly ignored (Jones *et al* 2003). This is mainly because of the quite simplistic nature of the road collision grid reference point location itself, they are single points nested usually on a road network. This data coincides with other features on the road network such as speed cameras, cycle lane length and pedestrian crossings. However other data linked to the collision hotspots include school location, weather, and light conditions which cannot be extrinsically linked to collision location; their attributes are represented at another scale.

It is difficult to quantify in road collision analysis how much the causal factors of a road collision differ from scale to scale. For example, a hotspot may have occurred at a junction in a borough in Outer London. The hotspot will spatially lie at a junction and typically entail a collision between two cars. The other characteristics from the collision data show that there is a cross section of people involved in terms of age and gender. Initially it would be apparent it is a hotspot because it is at a junction with a high number of collisions and there is potential for increased vehicle interaction and therefore collisions. However it is difficult to depict the causal factors which may be not apparent/openly visible. For example it could be a route way or it may be that there is a pattern with the type of people and their socio economic position and an inconsistency in their driving or lack of concentration. In short, the causal factor may be on a larger scale, the time of day is not associated with a spatial scale or location and plays a prominent role in being one of the causal factors of road collisions.

3.12 Ranking hazardous locations

The ranking of hazardous locations or hotspots has been a contentious issue over the years in the domain of road safety. Ranking of hotspots is primarily attributable to the challenges placed on funding for road safety management, and by ranking hotspots the most 'high risk' are dealt with first. This study uses ranking as a procedure to select the most spatially dense areas of collisions: however in the literature and in practical terms there are significant variations as to what method of ranking is best. This section seeks to highlight the different types of ranking available and the how this study determines the most high risk areas.

Nearly all road safety practitioners use GIS in order to determine hotspots, using perhaps thresholds per kilometre or the visual capabilities associated with GIS. However, increasingly, hotspots are determined using more sophisticated procedures, not just taking into account the collision frequency in a given spatial area or route but taking into account the severity and nature of the surrounding environment. Generally, hotspots are ranked by frequency; this study uses a spatial density measure based on frequency thereby adapting the previous applications which use a frequency measure and adding in a spatial dimension. Another key criterion which is often used to identify and rank hotspots is by severity. This study includes severity in the clustering process however it does not rank the hotspots, merely measuring the hotspots by density, and severity being on the variable. It is seen as more advantageous to try and prevent fatal collisions rather than property damage collisions. However the main drawback which it apparent in this study has been that fatal collisions are very rare and it is recognised that to try and relate them to each other across varying spatial environments is very difficult. John Adams (1995) highlights that a major problem faced in road safety analysis is that there are billions of *potentially* fatal events, but it is precisely because these events (or collisions) are recognised as potentially fatal that they are rarely so. Adams in 1995 determines that in 1985 London was the most dangerous jurisdiction in Britain with 759 road injuries per 100,000 populations (This stands at 426 per 100 000 in 2005). However company to this, London has one of the lowest recorded death rates ((7.3 per 100 000). Adams (1995) goes on to point out that the correlation between fatality rates and injury rates is very week. He asks the question whether London is the most dangerous or safest place with regards to road collisions? Another problem Adams (1995) acknowledges in the text is that generally there are not enough fatal collisions to form reliable statistical evidence. It would be easy to use a GIS to identify all the collisions which involved a pedestrian or cyclist or resulted in a certain level of injury: however if visual clusters are found this does not tell us the reason why the collision occurred in the first place or whether there are any other similar features to the collision in a certain area.

Another ranking method is to use collision rates. This method will be investigated in a subsequent chapter using the London road network and the collisions throughout it associated with the length of polyline represented creating a statistic of collisions per 100 metres. This in itself is unrepresentative of the spatial environment in which the collision took place as it 'averages out' the number of collisions along a certain length of road network. Another use of collision rates uses road traffic frequency and the number of collisions per million vehicles. Even ranking locations by collision rates can be problematic. A location on a low volume facility with only one collision could rank high because of the very low exposure. However, using limited funds to improve a location with only one collision may not be the most effective use of funds and may not prevent future collisions. Some agencies rank collisions by using a combination of frequency and collision rates.

Other methods used in many academic studies have focused on using empirical Bayes' method to rank hotspots (for a more detailed explanation see Chapter Five). The Bayes approach relies upon the comparison of frequencies and/or proportions of collisions at a given site with the amounts that would normally occur at similar sites. This study does not use Bayes or collision rates but rather adapts a spatial density method (based on collisions data) to identify the actual locations (using the $100m^2$ as the basic spatial unit) prior to selecting the collisions that fall into the hotspots. Each hotspot is given a density measure and a threshold is used, and all of the hotspots falling above a predefined density measure are used and their collisions subsequently selected and analysed to create the profile of the hotspot to be used in the clustering process. It is not until this stage that the attributes of the collisions will be depicted and patterns established between different hotspots which share similar collision attributes. There is no established correct procedure for ranking hotspots in the literature.

3.13 Conclusion

The spatial concepts which are associated with road collisions and the analysis and reduction of them are similar to those associated with many other geographical databases. However, the analysis of road collision data relies heavily upon quality and consistency in order to monitor and reduce road collisions. This chapter has tried to highlight that road collisions can be misleading and that by analysing data at single say country wide, level could conclude in very different results than local studies of specific networks, junctions or other basic spatial units. Road collision analysis constantly addresses boundary issues of how to constitute and define hotspots and the fundamental problems associated with this. Chapter Five addresses some of the more widely used hotspot identification methods whilst trying to maintain a level of uncertainty management in the process. The challenge that there is no geographical country wide or area wide guidelines associated with the determination of hotspots means that comparison between geographical areas remains somewhat difficult. This study aims to conjugate an analysis based on hotspot similarities rather than their actual geographical location and then to determine the best resource allocation mechanisms for accident reduction based upon observed similarities in accident attributes.

CHAPTER 4

PROFILING LONDON'S POPULATION WITH RESPECT TO ROAD COLLISION PROPENSITIES

4.1 Introduction

Analysis of the distribution of residential postcodes and areas of road collision casualties is not new (Haynes *et al* 2005). There have been many attempts to generalise the characteristics of the residential locations of people who are involved in collisions (see Abdalla *et al* 1997, Blatt *et al* 1998 and Lu *et al* 2000). However, these studies are limited; for example to general 'rural versus urban' profiling or to delineating social characteristics of an area of residence specifically focusing on deprivation. This chapter aims to further our understanding of high risk road users by profiling the residential postcodes of the casualties using geodemographics by means of the software Mosaic™ (Experian, Nottingham). This method offers a distinctive understanding of the patterns of consumer and lifestyle behaviour of the people most likely to be involved in collisions. Understanding which Mosaic types have higher propensities to be involved in road collisions based on the inputted variables will assist road safety experts in determining how to tackle these inequalities. Geodemographics has come to play an important role in public services, especially policing, education and health which has been driven by the need from central government for evidence based policy, improvements in local level spatial data infrastructures and a desire to develop a rational basis on which to set performance targets for public service delivery at a local level (Longley 2005). It has been acknowledged therefore that geodemographic profiling presents an opportunity to achieve savings by targeting communication programmes at populations to whom their message is most appropriate (Longley 2005). Although there have been no specific examples of using road collision victims and geodemographics, the literature suggests a strong association between socio economic variables and road collision propensity which will be discussed later in this chapter (see also Abdalla *et al* 1997). The innovative use of geodemographics in public service delivery shows a strong demand for better understanding of neighbourhood patterns, including road safety.

This chapter endeavours to understand and analyse the spatial propensity to being involved in a collision based on residential location – that is, assuming that where one lives affects the likelihood of where and how one is likely to experience a collision. The chapter begins with an outline of what geodemographics is including a critique and review of its literature. The following section describes two different methods for profiling and understanding the residential spatial risk of road users. Both methods use residential postcodes from the STATS19 database for the drivers and casualties.

Central London areas are divided into what are known as LONDON POSTCODES. Each small section of London is allocated a one or two letter prefix that corresponds to its compass location and then a following number to distinguish it from adjoining areas. For instance N3 is Finchley while the adjacent area is N2 which is East Finchley and they are both located in North London. The Centre of London is given the letter 'C' for CENTRAL, and is split into WC1 (where WC stands for West Central), WC2, EC1 (East Central), EC2, EC3, EC4. The EC1-4 postcodes are referred to as the 'City of London'. The W1 postcode is often thought of as central London because it is the tourist and clubbing centre, also known as the West End although it is actually slightly West of centre. Officially London centres on Charing Cross Station in WC1 (reasons for this will be outlined later).

This chapter begins by investigating the first stage of creating a hotspot typology whereby the residential location of the drivers and victims is profiled according to postcode data and the Mosaic classification. The overall hypothesis for this chapter is based on the premise that there are residential spatial patterns of involvement in road collisions within London based on postcode data. The basis for this assumption will be discussed in the next section, by investigating the relationship between collision involvement and the socio economic and environmental characteristics of the casualties involved in collisions.

The first method uses a vector based buffering approach to understand the changing population with regard to distance from central London and associated collision propensity indexes. This method uses an urban land use approach based on the different geodemographic types encountered with increasing distance from the centre of London. The second method uses the Mosaic types for the whole of London and profiles casualties and drivers, the results of which can then be applied to all the postcodes in London. This second method is supported by GIS based spatial analysis in order to identify high and low risk areas for collision involvement.

4.2 Socio economics and geodemographics

4.2.1 Overview

Traditionally most studies of road collisions have relied on collision statistics to address a range of safety related concerns such as the identification of road collision hotspots, the evaluation of safety programmes or the correction of irresponsible driver behaviour. However this section and the rest of the research challenges this assumption, arguing that using road collision data alone is insufficient to identify the main causes of road collisions. In most cases the cause of road collisions is not attributed to one single cause but it is an outcome of a complex process of interaction involving the driver, the vehicle, and the road environment. Therefore it is sometimes difficult to identify the main causes of a collision due to counts alone, which is why it is important to recognise that enrichment through use of geodemographic data can create a more accurate picture of who is more likely to be involved in a collision and possibly why. Pasquier *et al* (2002) indicate that collision frequencies segregated by location, time and type are generally low. They go on to argue that given the low rate of occurrence and the statistical nature of the problem the task of drawing statistically significant inferences by merely examining collision counts may not be an easy one (Pasquier *et al* 2002). However this just goes to exemplify the importance of using enrichment data in the form of geodemographic data.

It has been well documented that there is a significant and incontestable relationship between road collision involvement and social class (Klein 1980, Roberts *et al* 1996, Cooper *et al* 1998). In general most of the academic literature and government reports concerning road collisions and social and demographic indices are literature based, and do not provide robust empirical analysis of the current trends. In response to this narrow research base the main issue is that nearly all research in this domain is restricted to children, their socio-economic status and inferred road collision risk (Lawson *et al* 1991, Christie 1995). There has been only limited focus on understanding the risk propensity of adults within neighbourhoods and their different spatial socio economic risk in general rather than concentrating on one social variable such as deprivation or marital status. The literature suggests that there is a strong and incontestable relationship between pedestrian collision rates and social class and that the evidence is particularly marked for children. Yet this information leads us only to understand that children from lower socio-economic status have a higher risk of being involved in a road collision as a pedestrian and that they are more likely to be injured more severely. This incidence and risk is increased for children who belong to an ethnic minority (Transport Report, Vulnerable Road Users Report No. 19, 2002). There has been no comprehensive wide-reaching study that has analysed the effects of risk within different socio-economic neighbourhoods, that focuses on the population at large.

There have been a number of key reports and studies concentrating on the research about road collisions and the socio economic circumstances of the drivers and casualties (see Christie 1995, Abdalla 2001, Lawson 1990 and Haepers and Pocock 1993). However there has been an increased amount of research in recent years concerning child pedestrians. This emphasis arises because of the consistent government policies which focus on the reduction of the number of child fatalities, because of their high profile within society. Recent Department of Transport Reports on vulnerable road users have focused on child pedestrians as being of considerable risk of being involved in a road collision (Department of Transport, Vulnerable Road User Report 56, 2005). This section seeks to evaluate the various reports and research articles which have been written on the subject and how they relate to and complement this research. This part of the literature review is sub divided into sections and concentrates on addressing the specific nature and relationship of road collisions and socio economics. Section One will approach the research in terms of general factors relating to road collisions and socio economic variables. Section Two will focus on the research conducted on child pedestrians and social deprivation. Section Three will extend this evaluation further by looking at the relationships between children from ethnic minorities and road accident involvement. Section Four will highlight other research that has been conducted in an attempt to identify high risk road user groups for example the older driver and the young male driver. It is worth mentioning that most of the research studies I will be evaluating are UK based: however there is an international literature within this field, some of which I will be using to highlight that this phenomenon is not limited to the UK. Section five will outline the main limitations of these studies and various recommendations in terms of what my research hopes to bring to this field and areas for future work.

4.2.2 Road traffic collisions and socio economic variables

The majority of the research on this topic has focused upon Scotland, highlighted by a research article in 1997 concentrated on the Lothian Region (Abdalla *et al* 1997). This research focused on examining the relationships between areal social characteristics and road collision casualties (Abdalla *et al* 1997). Similar to the research design outlined in this chapter, Abdalla *et al* (1997) used postcode data for collision casualties and linked the data with 1991 Census data and the STATS19 records. The aim of the study is three fold focusing firstly on the distance between collision location and home address. Secondly the social characteristics of this area are then analysed followed by an analysis of the rates of casualties per head of population. This article supports the importance of linking the STATS19 data to socio economic data in order to make it more robust: however the main drawback within this report, that appears to be a common trend throughout the research, is the lack of attention paid to temporal variations and the lack of spatial context. It is important to understand how accident rates vary within a given type of neighbourhood because it will change over the course of a day, week, month and year. The research in Lothian Region used limited grouping within which to analyse the data. For example the study divided up the Census Output Areas in terms of fifteen most deprived and 15 most affluent based on only nine key variables from the Census. Abdalla *et al* (1997) highlight the concentrations of road collisions in deprived areas: however, there is no mention of temporal differences or trends. This research will extend this analysis by using neighbourhood classifications based not just on Census data but also Financial Spending surveys, behavioural information and sociological information. The research will use these neighbourhood classifications to create a more in-depth analysis of London society and the neighbourhoods that are more at risk than others.

4.2.3 Child pedestrian collisions and social deprivation

There are many different factors which can contribute to the increased road collision risk for children pedestrians of lower social classes. The majority of the studies are small scale, with no national policies having been published. The most comprehensive report in the UK on this subject was conducted by White, Barker and Raeside (2000), and concerned the relationship between road traffic collisions and children from disadvantaged areas. The main findings of this report can be summarised as:

- road traffic collision risk for child pedestrians is class related
- injuries of child pedestrians involved in collisions in lower socio economic areas are more severe than those that occur in higher socio economic areas
- Children of single mothers are twice as likely to be involved in a road collision as pedestrians than children living with two parents
- There are significant ethnic differences (to be discussed in greater detail in the next section)
- Anti social and over active children are more likely to be involved in a road collision
- Child pedestrian collisions during journeys to and from school are more common in low socio economic status areas than in more affluent areas

In this report many possible influencing factors were identified as increasing the risk of child pedestrian collisions. These factors are determinants of social exclusion including variables such as unemployment levels, low incomes, poor housing, high crime environments, bad health and family breakdown. Generally children from low socio economic backgrounds have a greater exposure to hazards that may result in a higher risk of road collisions (Christie 1995). It is clear however that few studies have sought to address the issues of family factors with reference to the relative affluence or deprivation of the area in which the household resides. As mentioned Judge and Benzeval (1993) conclude in their study that children of lone mothers have the highest death rates of all social groupings and that lone parenthood is a risk factor in relation to traffic related deaths. This is possibly because they are left unsupervised more often.

The literature suggests that there is a strong relationship between low socio economic status, ethnic minority membership and road collisions. This relationship will be discussed further in the next section. The main drawback in this area of research is that the standard accident collection procedure, the STASTS19 records, does not collect information regarding ethnic status. A study by Lawson and Edwards in 1991 indicated that Asian children are over represented among pedestrian fatalities.

There has been continuing research regarding the risk of child pedestrian collisions and the physical environment. Most of the research has been focused on the area level, it is apparent there is a relationship between city planning and road safety among children. Christie (1995) suggests that the layout of residential environments influences the safety of child pedestrians. A study in Australia by Robinson and Noland (1997) concluded that over 79% of road collisions involving child pedestrians occurred in a driveway, carport or garage of home addresses. Overall there have been many studies into social deprivation, locational factors and road collision involvement; however work by MacIntyre *et al* (1993) highlights that there has been limited research into variations of socio economic and cultural features of areas which influence health and the likelihood of death, particularly in a road collision. However there is conflicting evidence as to whether people of low socio economic status have poorer health due to areas in which they live in being health damaging or whether ill health and mortality is wholly explained by personal or socio economic factors.

Christie (1995) surmises that children from lower socio economic backgrounds are more likely to be involved in pedestrian related collisions, because their activities involve higher rates of risk than those of their counterparts from higher socio economic backgrounds. These activities are also more likely to be unsupervised and to take place in unprotected environments, namely the streets where the children reside. This report highlighted the distinctive relationship between social class and risk of death of child pedestrians (this research was conducted at a household level). Christie concluded that children from the lowest socio economic group are up to four times more likely to be involved in a road collision as a pedestrian than their counterparts in higher socio economic groups (Christie 1995).

The most recent work in this area is highlighted in two papers from the Centre for Transport Studies at London's Imperial College (Glaister *et al* 2002, Graham *et al* 2003 and Graham *et al* 2004). The first of these papers looks at whether the level of socio economic wellbeing influences child collision rates (Graham *et al* 2004). The approach differs from previous studies in this area, whereby instead of taking the socio economic status of the victim, they take a small area based approach of the pedestrian collisions. The aim is to ascertain the relationship between deprivation and casualties and is not straightforward: for example it may be that deprivation is more commonly found in dense urban areas and collisions occur more frequently in high population density urban areas. In short, areas of similar density but different levels of deprivation suffer similar casualty numbers or it could be that areas of similar density have very different casualty numbers depending on levels of deprivation. The article seeks to disentangle the effects of 'area' from the influences of socio economic characteristics. The study is UK wide and uses STATS19 data and 1991 Census data by ward. Information is therefore aggregated at the ward level. It concluded that deprivation is only one of many factors that will influence the number of child pedestrian accidents. Other factors that will influence involvement are listed below (Glaister *et al* 2002):

- Absolute number of children in a given area
- Volume of traffic flows
- Physical nature of the environment
- Characteristics of the local road infrastructure
- 'Other' local specific factors

The complementary study by Noland and Quddus (2004) highlights a more general population picture of the UK in terms of socio economic status and road collisions. The article demonstrates that although there has been a significant reduction in road collisions in the past thirty years this has been specifically because of improved vehicle design, safety belt usage, detailed safety audits after crashes and comprehensive engineering measures. However some factors that have received very little attention have been the land use characteristics of an area, population densities and urban development in terms relating to collision rates. In this study, the data are aggregated to the ward level, with a general reference that wholly urban wards experience a lower level of fatalities compared to wholly rural wards. It builds on two main findings from the literature; firstly from a research study by Sawalha and Sayed (2001) indicating that commercial land use (in Canada) is associated with a higher frequency of crashes. The second fundamental research findings by Ossenbruggen *et al* (2001) which examines the location of shops and finds that typical shopping sites are more hazardous than village style shopping sites (generally because of lower vehicle speeds). This study concludes that urbanised areas are indeed more likely to have fewer accidents while in comparison areas with higher unemployment densities are more likely to have a greater number of road accidents.

4.2.4 The relationship between children from ethnic minority backgrounds and road collision involvement

This section relates strongly to the previous section, as there is an incontestable relationship between ethnic minorities and socio economic status (Christie 1995). Recent surveys by Christie (1995) whereby data were taken from Bradford, Bristol, London, Merthyr Tydfil and Reading conclude that children from an ethnic minority background have a disproportionately high pedestrian collision rate compared to their non ethnic minority peers. The main drawback of these conclusions however is the studies that have been conducted have been relatively small scale and therefore caution should be taken when commenting on the robustness of the studies. A subsequent limitation which manifests itself in the small scale studies is the challenge to disentangle ethnic factors from co-varying social and economic indicators.

Within the UK, it has been documented by Lawson (1990), Lawson and Edwards (1991), Haepers and Pocock (1993) that road collisions involving child pedestrians are over represented by children of Asian origin especially from the Indian sub continent. The main drawback for the analysis of this relationship is that no ethnic information is recorded in the STATS19 database. This makes it harder to delineate ethnic trends in road collision occurrence. The first unpublished report to draw attention to this relationship was Lee (1986) whose report in Rochdale concluded that Asian children under 16 were twice as likely to be involved in a road collision; however the research failed to stipulate a differentiation between collision 'types'. The second influential report to be written was by Lawson in 1990 (updated in 1991 by Lawson and Edwards) whose study based in Greater Birmingham found inconclusive evidence that although there was an ethnic bias for children under 10 to be involved in a road collision, the trend reversed for children aged 11-14. Finally the most recent report to be published was by Christie (1995) who confirmed the over representation of non white ethnic children in road accidents: however, this study failed to identify whether any particular ethnic minorities were at high risk.

There are two factors which have been identified in the literature that influence ethnic minority involvement in road collisions (Christie 1995):

1. Ethnic minorities are less adapted to the UK traffic environment
2. Social deprivation

The association between social class and collision risk is incontestable. A number of studies have focused on finding which socio economic variables have correlated with collision rates (Chichester *et al* 1998). In general, the main issue is whether the observed relationship between ethnicity and collision involvement reduces to the same socio economic influences that affect the majority culture or whether there are cultural influences on exposure, supervision, socialisation, and education opportunities over and above those that follow from minority socio economic status.

4.3 Geodemographics

4.3.1 Overview

The premise of geodemographics is outlined by Flowerdew and Leventhal (1998) that 'birds of a feather flock together'. This phrase rests on Tobler's First Law of Geography whereby in 1970, Waldo Tobler expressed:

> *'everything is related to everything else but near things are more related than distant things'*
>
> Tobler (1970)

Both these ideas converge to form the idea that there is some form of neighbourhood effect. Demography is the study of the characteristics of human populations, thus geodemographics can be described as the study of population types and their geography. Geodemographics is largely concerned with the analysis of local scale neighbourhood geography. However, there is an important discussion as to the difference between small scale population analysis and larger scale area wide analysis. Therefore the question is begged: "To generalise or not to generalise". There are arguments for and against this idea. For example 'there is a need to recognise that cities are increasingly intersections of multiple webs of economic and social life, many of which do not interconnect' (Knox and Pinch 2000). However at the same time, Knox and Pinch's original emphasis that declares that 'we cannot generalise about the city' does not imply that local patterns of a given attribute should not be compared or explored. It is clear from the literature, however, that the 'neighbourhood effect' does exist; Tobler's law is still relevant to the study of the theory behind geodemographics. Geodemographic systems use a variety of different types of data, but the most important is national census data, and by applying a clustering algorithm to these data, composite variables are created that are deemed to be related and create the final typology.

4.3.2 Brief history and context of the current geodemographics phenomenon

In the 1920s, urban geographers were specifically interested in classifying the urban landscape. The motivation behind the new paradigm of research was to unravel the complex relationships within the cityscape of social, demographic and economic data. The result of this shift was the study of small areas within the city and the adoption of meaningful area typologies based on the mix of available data. The evolution begins with the research conducted by Park and Burgess in Chicago in the 1920s (Batey and Brown 1995). These urban sociologists used empirical urban data to develop and test theories about urban form and structure. A second wave of interest was seen in Los Angeles and San Francisco in the 1950s led by Shevky. The research adopted was known as 'social area analysis' (Batey and Brown 1995) and drew upon data attributes such as economic status, family and ethnic status. This new more robust type of analysis had several meaningful initial functions which ranged from delineating socially homogenous sub areas of the city and comparing two different points within the city and generally further the field of research within this expanding domain.

In comparison, the UK's advancement in area classification was restricted because of the lack of small area census data. It was only by the mid 1950s that the first small area classification breakthrough came about in the UK. It was the City of Oxford which served as the prototype for establishing small area census tracts, which however were still a lot bigger than the ones established in the USA. It was at the London School of Economics (LSE) where the most important breakthrough was to come about. The two researchers Moser and Scott developed a more weighted approach to method, scale and influence. The momentum of research gathered and spread to various local authorities in London and to Liverpool City Council and Manchester. During the 1950s and 1960s some preliminary research provided the basis for the conception of Richard Webber's ACORN (A Classification of Residential Neighbourhoods) typology. Webber, at the Centre for Environmental Studies in the 1970s created this typology for wards and parishes to fundamentally examine consumer behaviour within these small areas. A further shift came about when Webber moved to CACI, the only commercial organisation analysing small area variables. ACORN was created using the 1981 Census and initially dominated the market place. It is important to remember that these initial applications were all focused on the private sector, using the classifications to target consumers for sales marketing. There was little or no research into the use of geodemographics for the public sector.

In 1986, the software that is used for this research study was released. Created by Richard Webber as a rival to ACORN, it was called Mosaic. Mosaic is distinctive geodemographic software in that it is based upon a classification of postcode-based areal units instead of enumeration districts to originally create 58 neighbourhood types. For this original classification 54 variables were used (compared with the 2001 Mosaic where over 300 variables were used). By the 1990s there appeared an increased need for more complex data solutions for the end users and what was needed in terms of data variables. Brown (1991) defined geodemographics as:

a shorthand label for both the development and the application of area typologies that have proved to be powerful discriminators of consumer behaviour and aids to 'market analysis'

Brown (1991)

4.3.3 Geodemographics of the 1990s and 2000

The most important dataset within the classification process remains the Census data with the most recent census in 2001 creating a surge of new classifications such as Experian's Mosaic (2004). Experian has the advantage when creating Mosaic of having access to a wide variety of data sources as it is the UK's largest originator and owner of consumer data. Just over four hundred variables were used to create the current version of Mosaic based on the publication of the 2001 UK Census. Fifty four percent of the data used to build Mosaic is sourced from the 2001 UK Census. The other forty six percent comes from Experian's own Consumer Segmentation Database, including information about the Electoral Roll, Shareholders Register, House Price and Council Tax information and ONS local area statistics. Figure 4.1 shows the types of data used to build Mosaic.

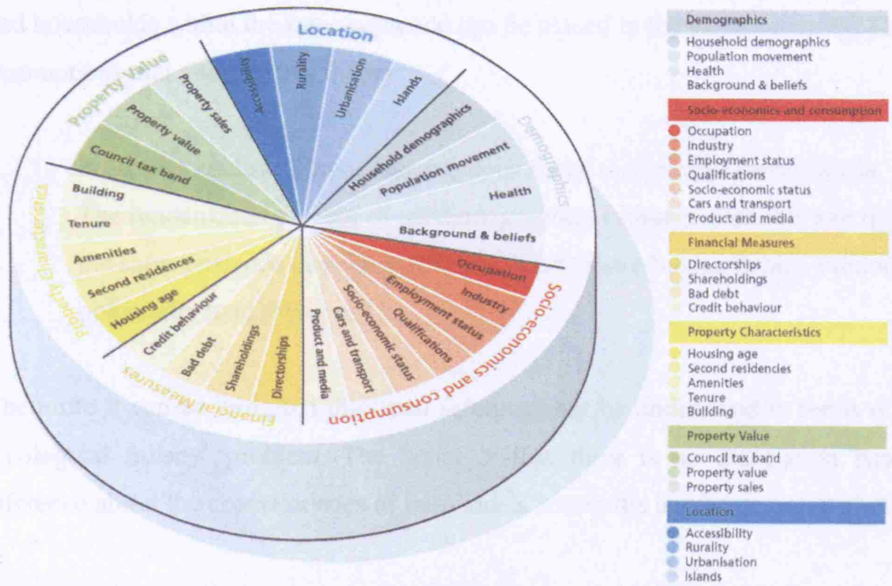


Figure 4.1: *Mosaic UK data sources (source: www.business-strategies.co.uk 2006)*

4.3.4 A critique of geodemographics

One of the main sources of criticism of geodemographics is asking the question, how accurate is it? Academics, especially geographers are sceptical of the process of creating these typologies, the methods of clustering and managing the error that occurs. It is true that not every postcode area can consist of all the same 'types' of people. As Voas and Williamson ask: 'How many different ways are there of being different?' (Voas and Williamson 2001). It is has been acknowledged that geodemographic classification, whereby small areas are grouped together into a number of categories depends on there being enough commonality to make such a typology meaningful. Are geodemographic classification techniques ineffective? Harris argues that the answer to this question depends on the user requirements of the geodemographic classifications (for example: commercial use or public service use), in response to Voas and Williamson's (2001) suggestion of 'the near – impossibility of defining a good general purpose classification' (Voas and Williamson 2001). It can be argued that the rapid rise of geodemographics and its continuing wide ranging applications is significant proof that it is useful and applicable in today's society.

It is not to be neglected in this research the effects by which individual persons might be misrepresented by the MOSAIC typologies (or any other geodemographic or lifestyle typologies). The theory of geodemographics is based on the fact that similar people tend to cluster together and households within the same postcode can be placed in the same category. There are two basic assumptions that underlie this theory

1. Postcodes enclose a broad homogenous social – economic environment
2. The fundamental precept of geodemographics is that households and individuals living in the same postcode or neighbourhood tend to share buying habits, product preferences and potentially road safety risk.

Therefore it can be criticised that road safety cannot be understood in terms of areas due to the 'ecological fallacy' problem. The belief is that there is an association based on fallacious inference about the characteristics of individuals within the aggregate population.

The ecological fallacy refers to the principle in empirically based social research that it may be illegitimate to make generalisations from data obtained between different settings, whether by aggregating data or by disaggregating it. This does not mean that identifying associations between aggregate figures is intrinsically defective, which would invalidate much economic analysis; nor does it mean that no inferences can be drawn about associations between the characteristics of an aggregate population and the characteristics of groups or smaller units within the population, which would make sampling impossible. What it does say is that the process of aggregating or disaggregating data may conceal the influence of disparate variables in different settings.

The ecological fallacy can be understood in making an assumption that any specific individual shares the general characteristics of his or her neighbourhood and its population. The more artificial the neighbourhood unit, which in this instance is the postcode the greater the risk of ecological fallacy because these formal administrative boundaries neglect the natural of homogenous groupings of the population. The risk of committing the ecological fallacy is greatest in areas of population diversity and so varies geographically. The severity of risk needs to be judged with respect to the context and aims of the analysis.

From the Census in 2001, there have been many studies in London analysing the demographic diversity within London (give examples). Specifically, a DMAG Briefing 2006/26 by the Data Management and Analysis Group for the Greater London Authority published information regarding Londoners who have English as a second language. For example 23 per cent of Londoners aged 16-34 have a first language other than English. However these statistics do not give a geographical disaggregation at a smaller scale within London to indicate diversity. It can be argued however that postcode granularity itself actually minimises the risk of ecological fallacy, certainly more than if a larger scale administrative geography were to be used such as output areas. There are 1.6 million postcodes in the UK which cover approximately 15 households. This granularity although is based on administrative boundaries, it is the smallest geographical boundary in the UK. Census Output Areas are built using postcodes as their building blocks and comprise approximately of 125 households. Therefore, although the ecological fallacy is indeed present in any geodemographical research, by using postcode information there is a minimisation in the presumptions associated with the ecological fallacy.

4.3.5 Methodological criticism

In order to fully understand the application of geodemographics to the concept of road safety it is important to embrace the criticisms and notion of error. This is not intended to undermine the relevance of the application just merely be aware of the pitfalls. It is true that over the years geographers have been sceptical of geodemographics. This section is intended to outline these criticisms and attempts to explain what this means for this research.

As with GIS, the use of geodemographics entirely depends of the user requirements. Geodemographics relies on the concept of the classification of small areas, which leads to the question 'How different are small areas from one another?' (Voas and Williamson 2001) The main argument surrounding geodemographics concerns the subjective emphasis of the clustering techniques. Clarke and Birkin (1998) comment that cluster analysis is a very subjective operation, involving important decisions on the user's behalf (Birkin *et al* 1998). This can range from the number of variables used, the clustering method, the number of clusters and how to manage poorly classified cases. Due to the advancement in this area, many companies now use non census data: for example the 2004 Mosaic uses a wide range of data from MORI financial surveys to electoral roll information and lifestyle data. This in itself gives the geodemographic software different weightings in terms of increased accuracy and certainty because the number of variables used to create the clusters in increased. A more comprehensive argument has been put forward by Voas and Williamson (2001). Their scepticism lies in the fact that this classification is based only on 'official' statistics, and therefore it would be impossible to differentiate if two families entered the same returns whether one family was non smoking, generally short and vegetarian compared to smokers, carnivores and very tall. The argument here would be that is depends on the user requirements, and how many data were needed to profile people.

Voas and Williamson continue their argument by challenging that there are a 'modest' number of categories that the UK society can be divided into (2001). The conclusion of their independent analysis found that measures of socio economic disadvantage tended to cluster, especially the proportion of households with no car being the single best indicator of relative deprivation. For my research geodemographics is a tool which is intended to complement the breakdown of the data in order to understand the 'types' of road accident that occur in relation to different 'types' of people.

It is an important consideration that needs to be made in that all characteristics ascribed to individuals by their Mosaic Type should be noted that these characteristics are associated with the areas within which they reside rather than the individual residents that have these characteristics.

4.4 Geodemographics and road collisions: is there a relationship?

The use of geodemographics to analyse road safety is a recent innovation. Its influence is supported by research linking socio economic variables such as unemployment, low income, area of residence, educational level and road collision risk, race and marital status (for example Lawson 1990 and Haepers and Pocock 1993, Christie 1995, Kposowa *et al* 1998, Murray 1998, Abdalla 1999, Road Safety Report No 19 2001 Department of Transport). Most of the reports and research conducted in this field have been focused on children and only a handful of studies have bridged the notions of road collisions and geodemographics. The studies that have been carried out relate an aspect such as urban and rural differences to changing collision risk (see Blatt *et al* 1998 and Lu *et al* 2000).

Road collision analysis has been slow to acknowledge the relationship between residential area social characteristics and road collision drivers and casualties. Social class as a discriminator for road collision risk has been addressed only by a minority of research papers (see Hasselberg *et al* 2005; Laflamme 2005). Research in Scotland (see Abdalla 1997 and Abdalla *et al* 1997) has considered deprivation indicators[1] from the 1991 Scottish Census as an indicator for road collision involvement. One of the key findings concluded that child casualties who came from families in social classes IV or V (semi skilled or unskilled jobs) were overrepresented in the total number of child casualties (Abdalla 1997). The influence and effect of certain residential layouts and housing types has also been found to cause an overrepresentation in collisions involving children (Christie 1995). Furthermore research undertaken by Hasselberg *et al* (2005) presents results for Swedish young adults that show that drivers with a basic and secondary education experience a greater risk of crashes of all types than drivers with an experience of higher education. In addition, the study found children of manual workers showed a 60% greater risk of being involved in any type of collision. These findings support the potential use of geodemographics as being a good indictor for understanding the 'who' and the 'where' of the people experiencing increased road user risk.

Two leading commercial geodemographic providers dominate the UK markets, Experian Ltd (Mosaic) and CACI Ltd (A Classification of Residential Neighbourhoods: ACORN). For this analysis Mosaic will be used to categorise the unit postcodes (of the drivers and casualties) into neighbourhood types. These types are based on social and demographic proximity and built

---

[1] Variables included; proportion of unemployed people, proportion of people with no car, proportion of people of pensionable age, proportion of people in a lower social class.

environment characteristics. Geodemographic classifiers cluster small areas on the basis of social similarity rather than locational proximity (Webber and Longley 2003). The core of this paper lies in the relationship between geodemographic attributes used to create the neighbourhood types and how they can assist the profiling of high risk road users. Mosaic classifies 1.6 million British unit postcodes into 61 'lifestyle' types. These types describe socio-cultural and socio-economic behaviour. There are more than 350 variables taken from sources such as the 2001 Census, Family Expenditure Survey, MORI's financial surveys and Experian Lifestyle Surveys. These data are used in statistical cluster analysis to build the 61 neighbourhood types which can be aggregated to 11 Mosaic groups.

Existing approaches to understanding road user risk in area social terms have been centred on using Census data, specifically deprivation indicators to determine a relationship between those people who experience high levels of deprivation and their overrepresentation in road collision statistics. This thesis uses geodemographics instead of Census data primarily because of the large potential geodemographics offers in terms of the wide ranging data sources which are included in the cluster analysis. Using geodemographics for road collision research enables the user not only to create a more succinct profile of the high risk user but also to target reduction strategies more effectively because of the inclusion of information regarding the most commonly used media outlets and preferred retail chains used by each Mosaic Type.

Findings in a recent paper by Webber (2004) suggest that neighbourhood effects such as income profiles, consumer behaviour, social grade and marital status are present at a range of scales. In a correlation matrix in this paper there was evidence that behaviours for which Mosaic Type is a good discriminator tend to be the same behaviours for which social grade, tenure, terminal education age and income are also powerful discriminators. In contrast to this finding, the behaviours for which Mosaic is not a powerful discriminator tend to be ones for which marital status, age or gender are relatively poor discriminator and vice versa. Therefore neighbourhoods are more homogenous in respect of status than life stage (Webber 2004). These neighbourhood effects cover a wider geographical extent than just a street or postcode, and geodemographics incorporates these contextual effects in ways which are akin to multi level modelling.

In a wider context, since 1997 there has been a renewed interest in academia and government in the use of neighbourhood classifications (Longley 2005). In policy terms, these developments have arisen from the opportunity to improve efficiency by targeting preventative communication programmes towards those most at risk (Longley 2005). In recent years, these programmes have centred on policing and health needs (see Ashby & Longley 2005), and with these public service applications comes the opportunity and methodological feasibility to apply geodemographics to road safety research. In response to the narrow research base is the issue that nearly all research in this domain is restricted to children and their socio-economic risk as been shown in the previous discussions. There have been a limited number of studies which have aimed to explore understanding the risks faced by adults within neighbourhoods and what can be deemed their 'risk exposure'.

Hauer (1980) gives a formal definition of exposure (related to the risk of a collision) as follows:

'A unit of exposure corresponds to a [probabilistic] trial. The result of such a trial is the occurrence or non occurrence of an accident (by type, severity etc). The chance set up is the transportation system (physical fatalities, users and the environment) which is being examined, and the risk is the probability (chance of an accident occurrence in a trial) and this describes the safety property of the transportation system examined'

Thus the ideal measure of exposure is one which is closely related to the opportunity of a road collision i.e. exposure is 'a condition which must be present in order to have an accident' (Tobey *et al* 1983).

Research by Julian *et al* (2002) in Paris stated that the majority of people who travelled on foot during the day were children, those not in paid work and the elderly, and she concluded that these pedestrians were at higher risk of being involved in a collision than other types of pedestrian. This study indicates that different levels of risk exposure do prevail between different groups in society, predominantly associated with mobility. Mobility and constraints on mobility have often been referred to with respect to the elderly and children. A person's mobility will in effect influence their exposure to traffic collision risk. Scheiner *et al* (2003) summarise that certain lifestyle groups (based on employment and income) have specific forms of mobility. Mobility here refers to 'short term' mobility (travel) rather than long term mobility (for example housing mobility) and in turn we can relate this mobility to differences in risk exposure.

From this section it appears that there is a clear need for a greater understanding of the effect of socio-economic factors as discriminators of road collision risk. However there is a need to progress to a more rounded conception of driver and casualty lifestyles in order to appreciate the nature of risk.

4.5 Method 1: The concentric circle method

4.5.1 Urban parameters and methodological guidelines

This method aims to investigate how people's risk based on their Mosaic Type changes with increased distance from a central London point (Charing Cross). Charing Cross was chosen as the centre for London because it is frequently used as the actual point in London for measuring the distance from London to say for example Birmingham or Brighton. Historically, it dates back to Edward I who installed a funeral cross for his wife at twelve locations from Lincoln to London, one of them being at Charing Cross. The methodology will identify, firstly, how a geodemographic classification of the total population (within each concentric ring) changes with distance from central London. The study then goes on to measure the total number of people involved in a collision aggregated by Mosaic Type set within the distance constraint from central London. The overall result is an indicator of the Mosaic types with the highest indexes indicating increased propensity to be involved in a collision (in terms of population for each concentric ring) and the corresponding risk index of being involved in a collision for each Mosaic Type. It is important to stress that this study is not analysing the location of the collision, but rather the residential location (postcode reference) of both the driver and the casualty involved in all collisions in Greater London from January 1999 to December 2003.

The growth of a city outwards will almost never be exactly concentric and even, and cities are usually organised into neighbourhoods to support educational and retail functions (Batty and Longley 1994). This methodology builds on the theory of Park and Burgess in the 1920s in the US city of Chicago. They built a concentric ring model which represented the different land use and socio economic areas of the city. Figure 4.2 illustrates the different concentric zones from the Central Business District. These concentric rings consist of population statistics towards the outskirts of the city. Although this is a largely out dated model, it provides a good theoretical grounding for why the residential population of London would change the further from Central London, and hence the risk propensities of the drivers and victims would change,

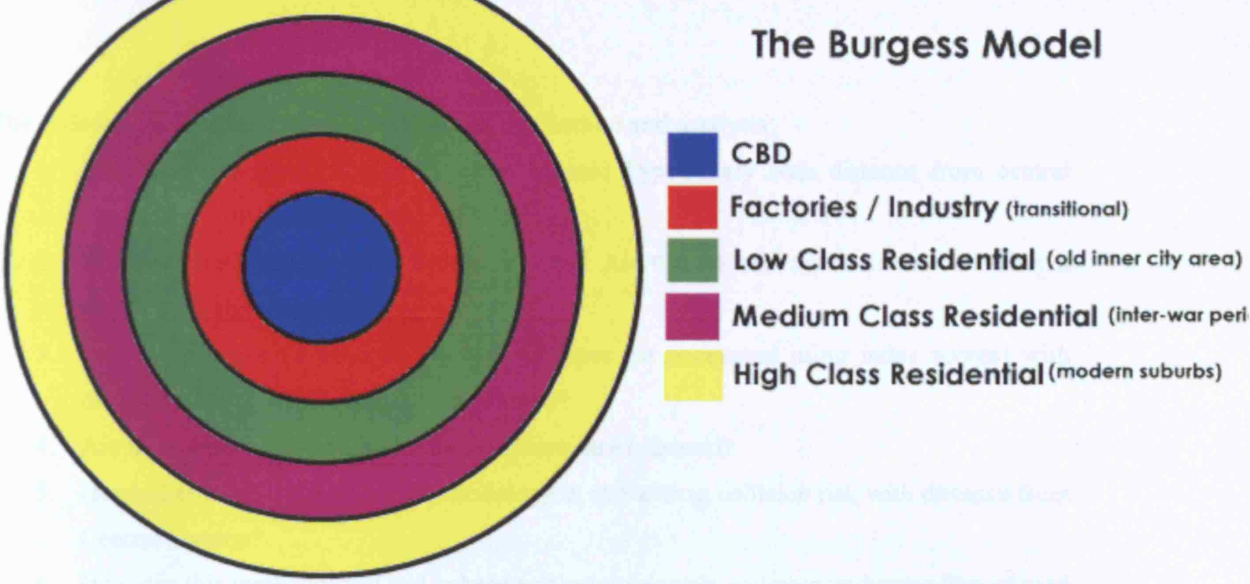


Figure 4.2: *The Burgess model (www.bennett.karoo.net/ topics/landuse.html May 2006)*

With this in mind, the concentric rings created for this study may overlook this neighbourhood effect of land use and infrastructure in relation to changing urban residential risk patterns. However, studies have neglected the link between distance from a central city point and the changing road user risk. The significant environmental and spatial factors which relate to the changing city attributes with distance from a city centre include changing land use and changing road network (usage and density) (see Anas *et al* 1998).

The effect of this urban form will have an impact on the Mosaic Types in each of the concentric rings around London. This impact will be largely reliant upon employment, transport and local infrastructure and the associated changing collision risks associated with these factors. Financial factors such as income often play the largest role in deciding where people live and as we have seen from the literature in previous sections of this chapter, being less financially mobile means a higher risk of being involved in a road collision primarily because of factors such as having a higher propensity to walk in order to complete a journey and a higher propensity to live in areas with high traffic volumes, and therefore an increased risk of being in the traffic environment as a victim. The importance of urban morphology and urban structure plays a vital role when measuring people's risk in terms of residential location and distance from central London.

The following questions act as a guideline for the method and analysis:

1. How does the proportion of different Mosaic Types vary with distance from central London?
2. What are the patterns of the Mosaic groups? Are the Mosaic types similar or different within each concentric ring?
3. Can the changes (if any) of the Mosaic types (as measured using index scores) with distance from Central London be explained?
4. Are there any anomalies, and if so, can these be explained?
5. How reliable are the data and methodology in explaining collision risk with distance from Central London?
6. How can this methodology and subsequent results enable a clearer understanding of road collision causation and risk exposure within London?

The differing risk exposure of the residents of London is likely to be unique (in terms of London's Mosaic population) because of its status as a world city and its characteristic social and economic structure. The results of this study will be influenced by the shape of London's urban growth and notably its sprawl (see Torrens *et al* 2000, Batty *et al* 2003). The concentric rings created around London and their subsequent values will be influenced by the distribution of population scattered within Central and Greater London. This effect will be discussed further in the results. A similar method using geodemographic scores and concentric rings can be found in Harris *et al* (2003).

One of the reasons why a single concentric zoning is appropriate for this methodology rather than one centred upon London's many different town centres is because the availability of population data for the concentric rings around London had to be uniform in distance (see Burgess concentric ring model) rather than a combination of many centres and many rings (based on the Hoyt urban model). A priori, one might envision that this arises because of the density of built infrastructure and work patterns, rather than the more incidental vagaries of neighbourhood structures. This concentric method also mirrors the transport pricing within London. Transport for London uses six concentric rings to calculate related travel costs via underground, bus and Docklands Light Railway.

For the purpose of this study, only spatial patterns of how road user risk varies with distance from Central London are examined. Temporal considerations have been excluded from the study (but their importance in collision occurrence cannot be stressed enough) because of time restrictions. Appreciation of the temporal aspect which contributes to road user risk (see Levine *et al* 1995 and Folkard 1997) should not be neglected and further studies will address this attribute when outlining patterns and reasons for risk.

4.5.2 Research design

4.5.2.1 Using a buffering tool

A GIS makes it possible to perform operations that are essential in decision analysis and decision-making: redistricting of boundaries, definition of buffer areas, and determination of the distance between objects. In redistricting, the boundaries of one territory can be modified or joined to those of another in order to form a new territory and to sum the values of constituent attributes. Buffering allows contiguous or non-contiguous territories or objects of different shapes and dimensions to be selected in order to form a virtual region or area without having to modify boundaries. Both redistricting and buffering capture the information on attributes of an area or region so that they can be managed or analyzed. Distance determination makes it possible to calculate the distance between two or more points on a map or the area of a territory.

To create the concentric rings or 'buffer zones' around Greater London, ESRI's ArcGIS was used as it has the capabilities to create easily modifiable buffers around a specific point. Charing Cross was chosen as the official central point of London (as suggested by Webber 2004 and as used in many cartographic maps of Greater London) and it grid reference was used as the point from which the buffers were determined. Using a buffering tool within ArcGIS, concentric rings were created around London at 3 mile intervals. Therefore in total there were 6 rings of the following distances:

- 0-3 miles
- 3-6 miles
- 6-9 miles

- 9-12 miles
- 12-15 miles
- 15-18 miles

The proposed stratum was set at 3 miles (28.3 Sq. Miles) in order to cover the whole of Greater London and enable a uniform comparison between the different buffer zones (i.e. at equal distances). It was proposed that a three mile radius around the centroid point provided coverage for specific inner London areas including the City and Kings Cross areas. However with this mind it is important to note that the area size of each buffer will be greater in area the further from Central London.

The data used to determine the collision victim location were obtained from the widely used road accident database STATS19, where information regarding many elements of the collision are recorded such as time of day, collision location, how many people were involved and what class they are (in terms of driver, passenger, pedestrian, cyclist). For this study a five year dataset for Greater London was used covering the years from 1998-2003: this included postcode data for both the driver and the casualty (disregarding whether the driver was injured in the collision or not). Each postcode for the driver and casualty was subsequently linked to a postcode point dataset which meant a point could be displayed on the map which represented a postcode (which represents approximately 15 households: Royal Mail 2005). For each concentric ring, the total number of postcode points was collated and a Mosaic Type was appended to each postcode.

4.5.2.2 Stats19 database: Preparation and explanation

Collision data for London are collected and maintained by the Metropolitan Police in what is known as the 'Stats19' database, which records all injury collisions within Greater London. This information is divided into three separate datasets over a five year time period (Jan 1998-June 2002) which include:

- Attendant circumstances (basic information about the collision)
- Casualty details
- Vehicle details

For purposes of this thesis the dataset was disaggregated by driver and casualty rather than using the collision location information from the Attendant Circumstances, just using the latter two datasets for this analysis (for this section of the analysis the road collision location is not being taken into account, as we are focusing on the locational residence of the driver and casualty involved in the collision). Both the driver and casualty datasets were kept separate for the purpose of maintaining the structure of the original data collection procedure and to avoid any confusion between the two datasets. The data consisted of only postcode data and Mosaic Type for each postcode (which has been appended to the dataset), crash reference and the easting and northing centroid point for the postcode in order to plot the residential location as point data in order to select the points within each concentric ring.

Within each buffer zone there are a number of postcodes which are incomplete which could not be identified because of incomplete Stats19 collision records. These records were omitted as an accurate Mosaic Type could not be appended to these data. These unclassified postcodes constitute roughly 1% in each buffer zone of the total postcode counts.

4.5.2.3 Establishing a risk index

Establishing a risk index for each concentric ring around London entailed identifying the base population and Mosaic Type counts, and counts of the collision victims Mosaic Types. The next step was to calculate the Mosaic count for each Type within each buffer zone, and the related Mosaic count for the drivers and the casualties. Below is the equation used to determine the index (a measure of differing scale with '100' being the expected or 'normal' value and if the number is below 100, the value is under represented and if is higher than 100 then the value is over represented).

1. Expected Value

$$\frac{\text{Total postcode count casualty/driver (buffer zone)}}{\text{Total base population households for buffer area}} \times \text{Individual Mosaic type total (eg A01 Global Connections) for each buffer}$$

2. Index value

$$\frac{\text{Casualty/driver total for each Mosaic Type}}{\text{Expected Value}} \times 100$$

In order to make the results meaningful, the ten highest percentages of total Mosaic household counts for London overall were used (representing a more meaningful representation of the population by discounting any Mosaic Types with low population percentages within London) These ten highest total Mosaic household counts can be found below in Table 4.1. The second column in Table 4.1 (below) relates to the total household population count for each Mosaic type in London. The third column relates to the percentage of this population which is in the buffer zone 0-3 miles around Charing Cross. The final two columns show the propensity of drivers and casualties (see ***4.5.2.3 Establishing a risk index***) living in that area that more likely to be involved in a collision.

4.5.3 Results

Mosaic Type	Total Mosaic population count for all London	0-3 mile %	Casualty Index	Driver Index
F36 Metro Multiculture	256971	36.16	129	126
E28 Counter Cultural Mix	140721	19.80	98	96
A01 Global Connections	133965	18.85	56	61
E29 City Adventurers	73677	10.37	79	89
D26 South Asian Industry	14874	2.09	108	110
D27 Settled Minorities	11066	1.56	189	181
A02 Cultural Leadership	8662	1.22	82	91
E30 New Urban Colonists	8053	1.13	86	103
E33 Town Gown Transition	7744	1.09	57	52
F39 Dignified Dependency	7675	1.08	119	88

Table 4.1: *Table of 0-3 mile buffer values for the ten most numerous Mosaic Types in London*

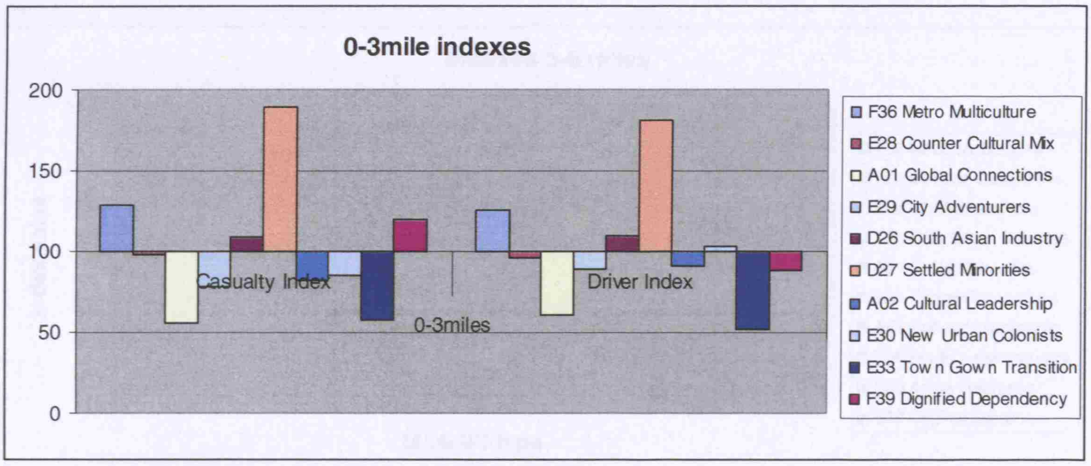


Figure 4.3: *Graph of index values of 0-3 mile buffer zone*

Mosaic type	Mosaic	3-6 mile %	Casualty	Driver
F36 Metro Multiculture	411870	22.94	109	104
E28 Counter Cultural Mix	379520	21.14	110	106
E29 City Adventurers	227194	12.65	82	86
E30 New Urban Colonists	196893	10.96	91	95
D27 Settled Minorities	196611	10.95	128	126
A01 Global Connections	144411	8.04	64	67
A02 Cultural Leadership	90620	5.05	72	83
D26 South Asian Industry	17773	0.99	69	68
C20 Asian Enterprise	16038	0.89	103	103
C19 Original Suburbs	14545	0.81	122	133

Table 4.2: *Table of index values for the ten most numerous* Mosaic *Types in the 3-6 mile buffer zone*

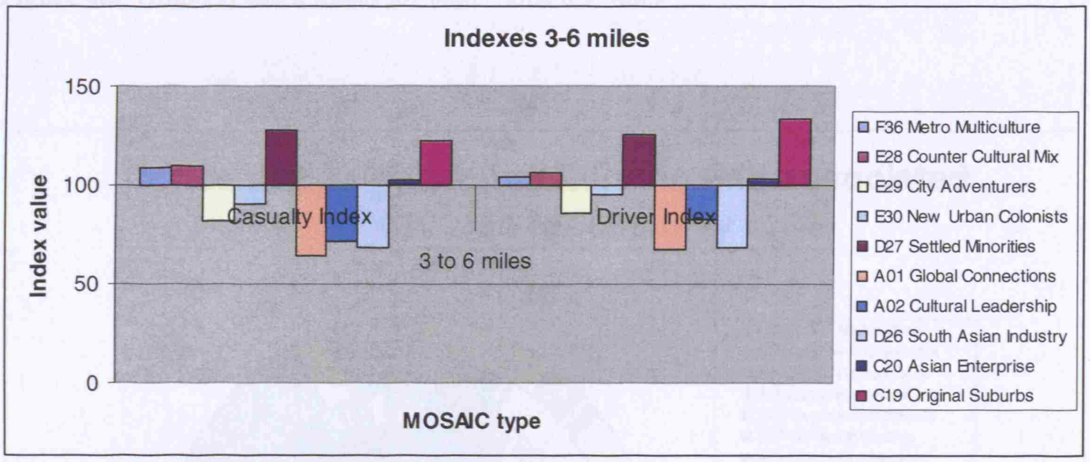


Figure 4.4: *Graph to show the index values for the 3-6 mile buffer*

Mosaic type	Mosaic count	6-9 mile %	Casualty	Driver
D27 Settled Minorities	488843	23.56	111	109
F36 Metro Multiculture	218624	10.54	115	102
C20 Asian Enterprise	179541	8.65	107	111
E30 New Urban Colonists	176332	8.50	75	80
C19 Original Suburbs	143555	6.92	104	112
A02 Cultural Leadership	141434	6.82	78	84
E28 Counter Cultural Mix	111087	5.35	100	105
H46 White Van Culture	101562	4.90	119	105
E29 City Adventurers	67439	3.25	68	74
D26 South Asian Industry	56324	2.71	110	114

Table 4.3: *Table of indexes for the ten most numerous* Mosaic *Types in the 6-9 mile buffer zone*

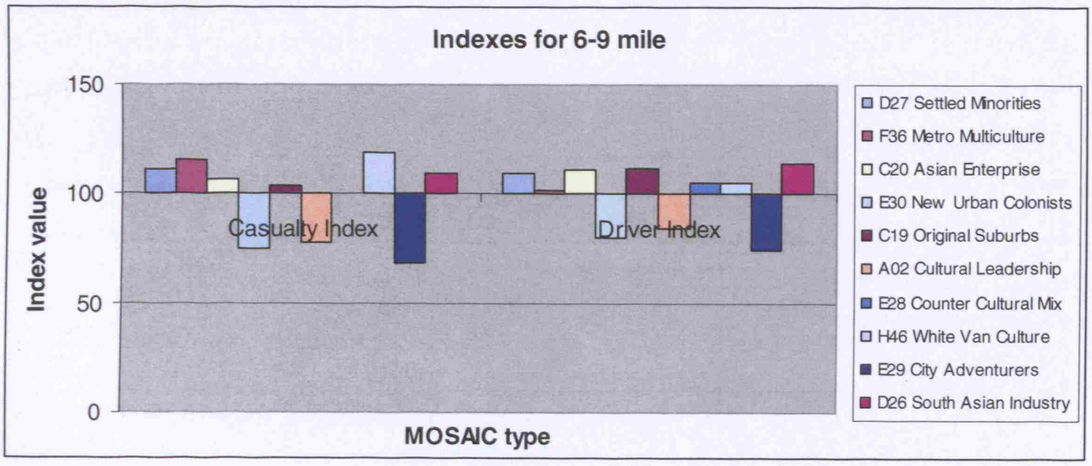


Figure 4.5: *Graph of index values for buffer zone 6-9 miles*

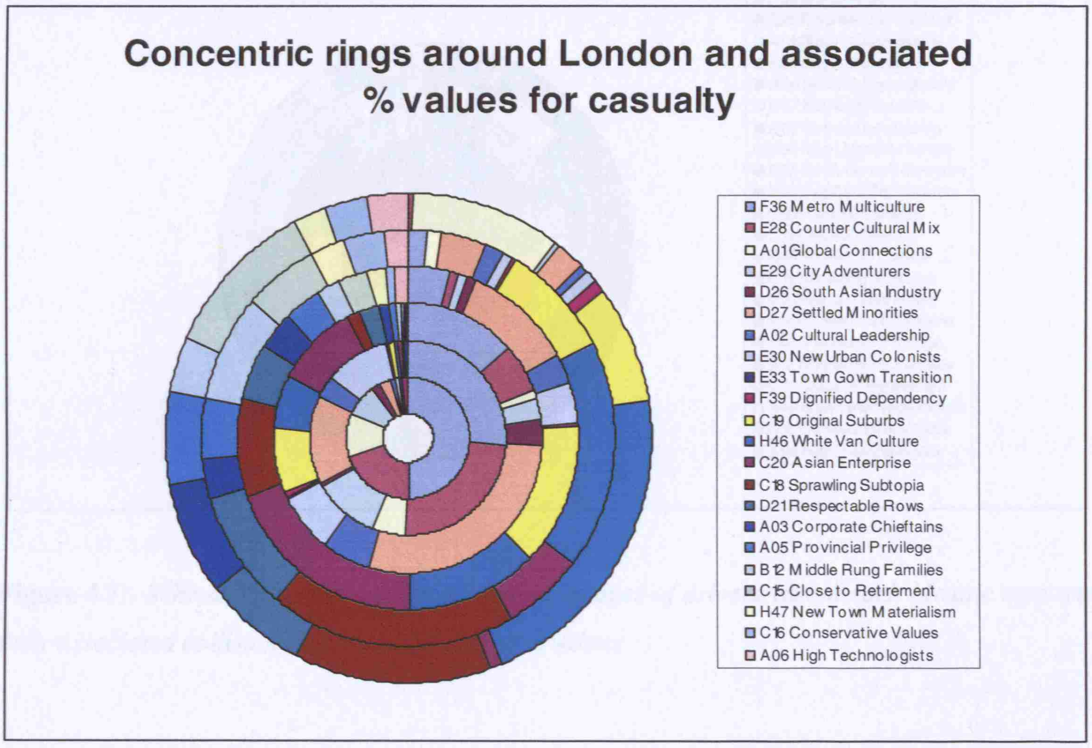


Figure 4.6: *Schematic diagram to show the percentages of casualties within each Mosaic type and their associated collision risk based on the index values. The types were selected if their percentage was above 1% of the population in the buffer zone*

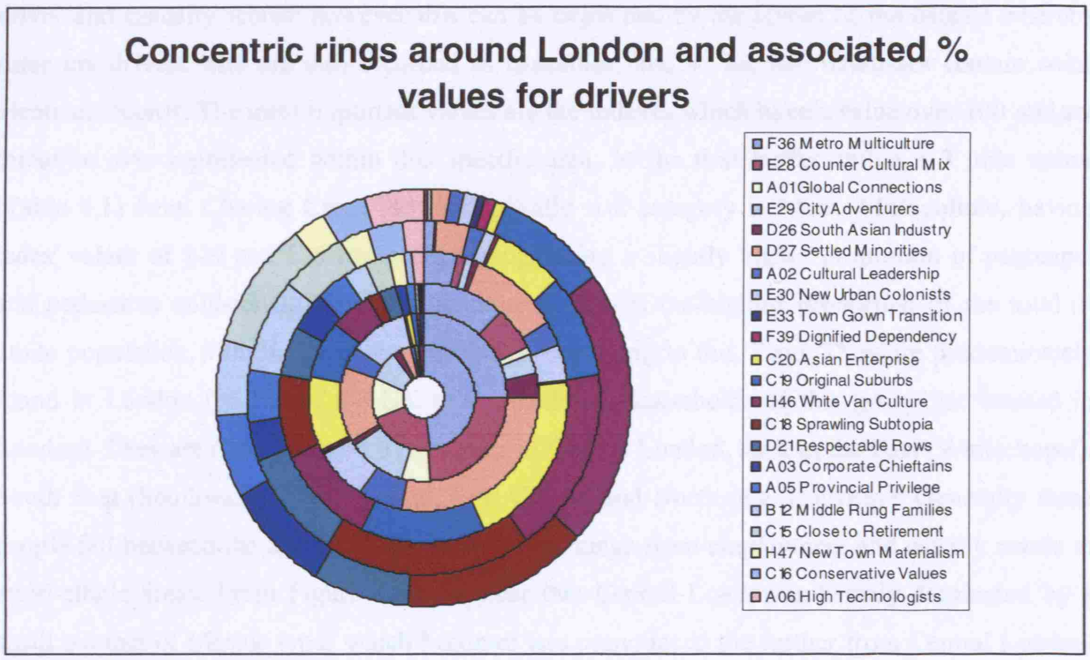


Figure 4.7: *Schematic diagram to show the percentages of drivers within each Mosaic type and their associated collision risk based on the index values*

This type of diagram (Figure 4.6 and 4.7) was chosen with the intention to visually display the percentages of Mosaic types within each buffer zone. It can be easily used to determine how the Mosaic types become less polarised in their risk with increasing distance from London's central point. All the rings consist of different area sizes; they do give an indication of the changing population at risk of being in a collision. Figure 4.7 shows clearly that F36 'Metro Multiculture' has a higher index score in the first 0-3 mile buffer than the 15-18 mile buffer, suggesting a geographical shift in geodemographic types and therefore risk indexes for this particular Mosaic Type. There is an important aspect of this methodology which will result from the increases in area size for each buffer which means that the area will possibly contain more Mosaic Types and therefore could become less polarised. However, with this in mind it is essential that a greater understanding is needed to explain the changes in Mosaic Types and those of which have a higher index score for being involved in a collision. This will be discussed in the next section.

4.5.4 Analysis and findings

The tables and figures in the previous section highlight a varied risk exposure index with increased distance from London's central point. It is important to note the similarities for both the driver and casualty scores: however this can be explained by the layout of the dataset whereby there are drivers who are also recorded as casualties, and so the two databases contain some identical records. The most important values are the indexes which have a value over 100 and are therefore over represented within that specific area. In the first buffer within a 3 mile radius (Table 4.1) from Charing Cross the most prolific risk category is Metro Multiculture, having index values of 129 and 126 respectively (suggesting a slightly higher proportion of passenger and pedestrian collisions). Metro Multiculture possesses the highest proportion of the total 0-3mile population with 36.6% of the population belonging to this Type. They are predominately found in London (96.1% of the UK total population households of this group are located in London). They are more likely to live in parts of Central London, such as the East (Whitechapel), South East (Southwark), West Central, East Central and North (e.g. Hackney). Generally these people fall between the age of 25-45, have low earnings from employment and usually reside in multi-ethnic areas. From Figure 4.6 it is clear that Central London is heavily dominated by a small number of Mosaic types which becomes less pronounced the further from Central London. Within 0-3 miles the second most dominant risk group is D27 Settled Minorities with a risk index of 189 (casualties) and 181 (drivers) respectively. Settled minorities feature within the first three zones from Central London, with an over represented risk index every time. Within London, they are more likely to be found in areas such as Tottenham or Walthamstow. As with Metro Multiculture, this Mosaic Type is highly multi ethnic, living in terrace housing with high levels of unemployment. There is a marginally higher risk index for casualties rather than drivers which suggests a greater risk as a pedestrian, cyclist or passenger in a vehicle.

By the second and third buffer (Tables 4.2 and 4.3), it is clear that the Mosaic type C19 'Original Suburbs' features highly in terms of an over represented risk index for all collisions. Suburbs are not city centre features, and this approximate distance from Central London has increasing suburban type areas, which are found in these buffers. That is not to say that the shape of these suburban areas is circular as the growth of a city can be very haphazard, merely that they can be found in these buffers. Their risk index is over represented within the 3-6 mile buffer with index values of 122 and 133 respectively. This Mosaic Type is generally found in areas such as High Barnet or Raynes Park. With incomes slightly higher and a more family orientated structure, a higher proportion of this Type owns cars and also uses public transport. By 6-9 miles, the risk index becomes more even, with more Types having an over represented index, but the indexes are above average measures but only just so. Mosaic Type H46 'White Van Culture' is absent as a high risk group in Central London, until 6-9 miles outside of London when type H46 has a casualty index score of 119 rising to 142 in zone 9-12 miles outside of London, and then dropping to an index score of 125 by 12-15 miles. 'White Van Culture' areas in Greater London include Borehamwood, Stevenage and Morden. In the 12-15 mile distance band, the highest risk index is C15 'Close to Retirement', with a considerably higher proportion of the total in the driver category.

With distance from the centre the risk indexes appear to reduce and the proportion of Mosaic Types becomes more evenly spread rather than polarised by a few Mosaic Types. Therefore it is necessary to include more than the top ten Mosaic Types based on household counts for the last two buffer zones. An example of this is Mosaic type E32 'Dinky Developments': this type falls into the top 15 in the buffer zone 12-15 miles and 15-18 miles. Within 12-15 miles, type E32 has a casualty risk index of 124 increasing to 139 in the next zone. The driver risk index increases from 131 to 157 respectively. Type E32 (the acronym 'Dinky' stands for 'Dual Income No Kids Yet') are to be found in areas such as Uxbridge, Croydon and Watford. They tend to live in terrace house cul-de-sacs, are typically moderately well off and a high proportion own cars.

Over-all, the results show the further out of London the less likely the patterns of risk are to be skewed. For example, it is clear that the high risk indexes nearer the urban centre are higher amongst the less well off, such as Mosaic Type F36 'Metro Multiculture' which is characterised by its deprivation and low socio economic status. This pattern mirrors the population distribution as a whole for that buffer zone. Mosaic types such as A01 'Global Connections' in the 0-3 miles buffer are actually highly under represented with a casualty index score of 56. The indexes of the first few buffer zones are extremely skewed with values ranging from 189 (Settled Minorities) to 56 (Global Connections). This skew becomes less pronounced with increasing radial distance from London, for example, in buffer 12-15 miles the range in risk propensity index score is between 76 (Provincial Privilege) and 125 (H46 White Van Culture). The risk indexes follow an consistent pattern with the only exceptions occurring if a more wide ranging collection of Mosaic types is analysed in the outer distance bands. This, as mentioned highlights differing and increasing risk indexes. Whether the risk index is high or low, there is a tendency to focus on the Mosaic Types which possess a high index score and therefore a higher likelihood of being involved in a collision. However the lower than expected scores are equally important (for example type A02 Cultural Leadership's index score drops significantly between zones 0-3 miles, where it is nearly 200, to the next zone 3-6 miles where the score is 60). These negative findings pose just as an important finding for policy makers than the positive scores.

4.6 Method 2: London index scores - a population profile of collision propensity

The analysis of road collision casualties concentrates on the people who reside in London and experience collisions within London as opposed to residents of London experiencing collisions outside of Greater London or people from outside of London being involved in collisions within London. Postcodes were excluded from outside this area in order to give a more accurate understanding of the risk faced by those that live in the capital. The aim was to elaborate the understanding of the nature of people's propensities to be involved in a collision based on geodemographic indicators and to assess the potential for reducing collision risk. Accordingly Mosaic geodemographic codes were appended to individual records of all the driver and casualty postcode data within London for the years 1998 – 2002. The first stage of the analysis entailed attaching each of approximately 100,000 postcodes for both drivers and casualties to a Mosaic Type. The geodemographic system was then used to analyse the incidence of drivers and casualties across the 61 Mosaic neighbourhood types. By standardising about an index value of 100 the Mosaic codes can be compared across London, thereby comparing different neighbourhoods or boroughs. As in the previous section, a value over 100 indicates a higher than average propensity to be involved in a collision while a score below 100 indicates a lower than average propensity to be involved in a collision. In the next section the postcodes have been appended to all the postcodes across London. These are then mapped using GIS to highlight the areas of high or low propensities. There are a few considerations which need to be addressed before the Mosaic patterns are outlined. There is no detailed outline of the typical vehicle ownership traits of each Mosaic Type and so inferences have been drawn from the Mosaic pen portraits regarding the types of cars (if any) that people are most likely to own and how they are likely to drive. If future work was to be done in this area, it would be beneficial to use the Census to depict car ownership. However due to time constraints this was not possible. For example, type A01 'Global Connections' is likely to have more than one car, and they are likely to be expensive brands. They are likely to have well paid jobs and value the safety of the people around them. Therefore it can be assumed that this Mosaic Type is likely to drive quite safely (as mentioned this is only conjecture). In comparison, F36 "Metro Multiculture', despite having low incomes, tend to own one car which they see as a status symbol. However their car driving behaviour would be different. [2]Firstly they would not use their car frequently because they would use public

[2] This is conjecture

transport to get around London and shop at local stores close by. [3]Secondly, they could possibly use the car for leisure, at the weekend, travelling to places which they are unfamiliar with.

4.6.1 Driver propensities

With regard to Mosaic Types and their population proportions in London, F36 'Metro Multiculture' has the highest proportion amounting to 12.32% of the total London population (Table 4.4).

Mosaic types	London %	Driver Index
F36 Metro Multiculture	12.32	94
D27 Settled Minorities	11.38	109
E28 Counter Cultural Mix	9.12	84
E30 New Urban Colonists	6.64	75
C20 Asian Enterprise	6.46	109
C19 Original Suburbs	6.31	101
E29 City Adventurers	5.84	65
A01 Global Connections	5.31	41
H46 White Van Culture	4.70	133
A02 Cultural Leadership	4.65	66
C18 Sprawling Subtopia	3.70	118
D21 Respectable Rows	2.54	122
A03 Corporate Chieftains	2.22	103
J52 Childfree Serenity	2.00	80
A05 Provincial Privilege	1.90	88
E32 Dinky Developments	1.25	132
C15 Close to Retirement	1.20	151
B12 Middle Rung Families	1.16	161
D26 South Asian Industry	1.16	108

Table 4.4: *Mosaic types and associated London population percentages and associated index scores for drivers*

However, people in F36 'Metro Multiculture' are less likely in relative terms to be involved in collisions within London, compared to type D27 'Settled Minorities' which has a slight over representation within London. Type D27 are characterised by living in Victorian and Edwardian two storey houses usually in the Inner City. Over time, this Type has been characterised by an outward migration to Outer London areas such as Edmonton as generational aspirations are

[3] This is conjecture

fulfilled. The population is fairly young and the infrastructure is usually run down. However there is a strong element of aspiration amongst this Type and they want to seen to driving expensive cars, with a high proportion driving elderly BMWs. Type C20 'Asian Enterprise' makes up 6.46% of the London population and has a slightly higher than average likelihood to be involved in collisions (they have an index score of 109). This type is fairly affluent, and this is reflected in the high number of expensive cars which are in the driveways of the inter war suburbs they inhabit -- areas such as Harrow, Wembley, Ealing and Hounslow. Type H46 'White Van Culture' has a significant over representation of being involved in a collision as a driver. This type is characterised by high levels of local geographic mobility because of the nature of their employment. They predominantly live in outer London suburbs around the M25. Their car ownership is high -- often two or more cars per household. In comparison type A01 'Global Connections' has a considerable under representation of being involved in a collision as a driver. This type is likely to drive expensive cars which can be chauffeur driven or they rely on taxis. This means that their exposure to the road environment is slightly less than most as they tend to live in Central London and being a passenger in a taxi means a safer environment because of the driver's experience. The cars that are owned are expensive and driven more and with more care and attention. Similarly type A02 'Cultural Leadership' also has a low collision involvement propensity. This type lives further out of London that type A01 in places like Ealing, Highgate and Dulwich. They live in what they describe as 'villages' and could contribute to the low collision involvement for many reasons, particularly the increased sense of neighbourhood and community in these areas. Whereby the overall environment and roads gives an appeal for safety and driving more carefully. Overall however there seems to be a general pattern of high Mosaic types which live in outer suburbs and have low income. Mobility is varied but the Types highlighted here have an over representation of being involved in a road collision and all have the use of at least one car.

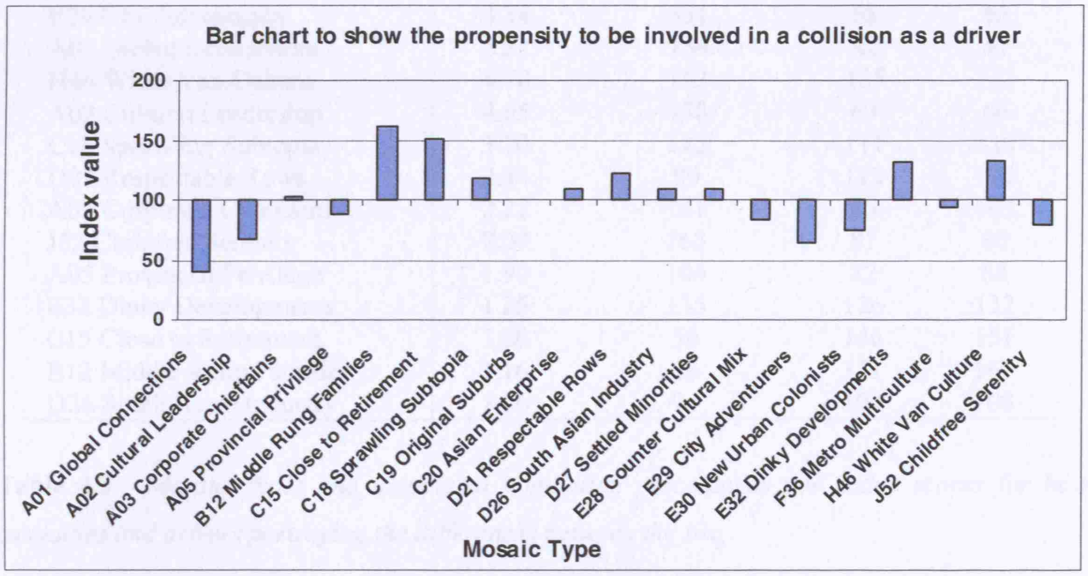


Figure 4.8: *Graph to show driver propensity to be involved in a road collision in London*

Figure 4.8 shows the index scores for all the Mosaic types with proportion of the population of London over 1%. Type B12 'Middle Rung Families' has the highest over representation as a driver in a collision, however they account only for a small proportion of the population. It is clear the Group A 'Symbols of Success' has a strong under representation as a driver in collisions. Both Groups B 'Happy Families' Group C 'Suburban Comfort' and Group D 'Ties of the Community' have Types which have a higher than average index.

4.6.2 Casualty propensities

The index scores for the casualties are calculated in the same way as the driver index scores. The reason for the two types of casualty is based on the organisation of the Stats19 data, whereby the data are divided into both casualty and driver details. In this instance it is important to be aware that if a driver were involved in a collision and were also injured he/she would be in both groups. The difference is that the casualty group contains details of pedestrians and cyclists, possibly indicating more vulnerable road user propensities.

Mosaic Type	London	London Index	Casualty	Driver
F36 Metro Multiculture	12.32	758	101	94
D27 Settled Minorities	11.38	680	113	109
E28 Counter Cultural Mix	9.12	751	91	84
E30 New Urban Colonists	6.64	519	77	75
C20 Asian Enterprise	6.46	508	105	109
C19 Original Suburbs	6.31	244	96	101

E29 City Adventurers	5.84	531	68	65
A01 Global Connections	5.31	734	42	41
H46 White Van Culture	4.70	143	135	133
A02 Cultural Leadership	4.65	438	63	66
C18 Sprawling Subtopia	3.70	112	111	118
D21 Respectable Rows	2.54	99	119	122
A03 Corporate Chieftains	2.22	161	86	103
J52 Childfree Serenity	2.00	163	81	80
A05 Provincial Privilege	1.90	104	82	88
E32 Dinky Developments	1.25	135	126	132
C15 Close to Retirement	1.20	38	146	151
B12 Middle Rung Families	1.16	38	151	161
D26 South Asian Industry	1.16	94	104	108

Table 4.5: *Mosaic Types and associated population percentages and index scores for both casualties and drivers portraying the differences between the two*

It is apparent from Table 4.5 that subtle differences exist between the driver and casualty collision propensities. For example Type F36 'Metro Multiculture' has a slightly above average representation as being a casualty in a road collision perhaps indicating a higher proportion of pedestrian risk within this Type. The same pattern also occurs for type D27 'Settled Minorities'. Predominantly the differences are minor and often the casualty index scores are slightly lower than the driver index scores.

Figure 4.9 below shows the Mosaic Types and their associated index scores. Type C15 'Close to retirement' has a higher than average propensity to be involved in a collision. This type is very different to the other high risk types. 'Close to retirement' has a high disposable income and they have a strong element of predictability and safety. It would be also possible to speculate that their reaction times might be slower. These people would live in Outer London in more satellite towns and therefore their mobility would be based on using the car especially to journey to work, however they would use their local amenities considerably.

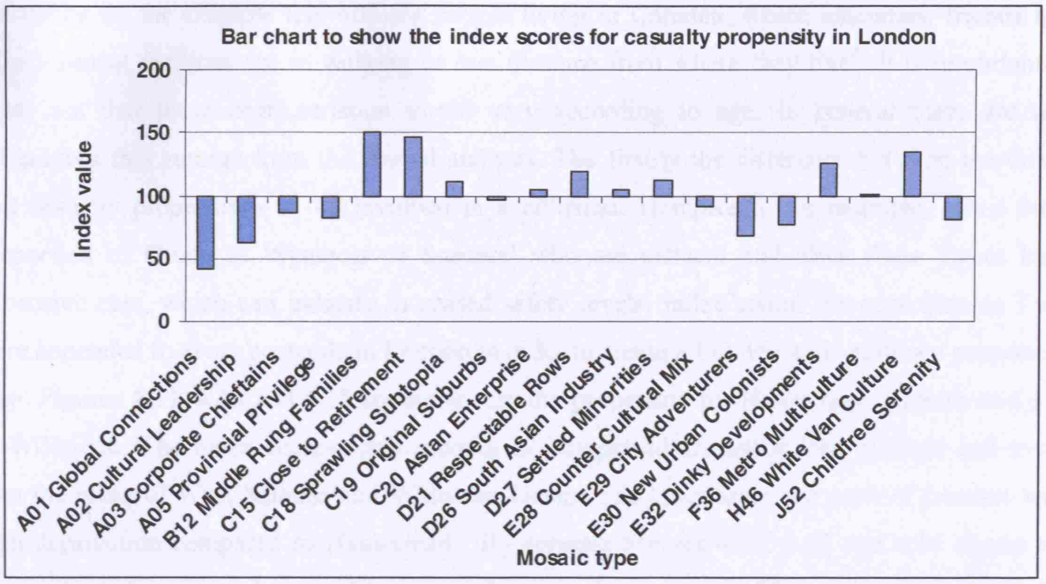


Figure 4.9: *Graph to show casualty propensity to be involved in a road collision in London*

There is generally a reduced propensity to be involved in a collision as either a casualty or a driver if one belongs to one of the higher income Mosaic Types such as those in Group A. Some of the over represented Mosaic Types involving pedestrians are often ethnic minority based; these include C20 'Asian Enterprise', D26 'South Asian Industry' and E28 'Counter Cultural Mix'. This supports studies linking ethnic minorities and their higher risk within the road environment (Report No. 19, Department for Transport 1999). This group in society, particularly the children, are seen as high risk in terms of being involved in a collision. In the next section, we will observe the spatial patterns within certain areas of London and the propensity to be involved in a collision based on postcode data. Preliminary observations show evidence that there is a clear pattern between deprived and affluent areas within London.

4.6.3 Spatial analysis of drivers and casualties in London and people's collision involvement propensities based on their postcodes

There are unambiguous spatial patterns within London with regard to susceptibility to being involved in a collision as either a driver or a casualty. There is an increased likelihood for people living in the outer London Boroughs to become involved in a collision as a driver, because these geodemographic types are more likely to own a car and drive to work, to the shops or to friends compared to, for example less affluent people living in Camden, where amenities, friends and entertainment facilities are in walking or bus distance from where they live[4]. It is important to point out that these characteristics would vary according to age. In general there are two differences that emerge from the spatial analysis. The first is the difference between the driver and casualty propensities to be involved in a collision. Hampstead, for example, has a large proportion of Group A 'Symbols of Success' who are affluent and often these Types have expensive cars, which can indicate increased safety levels. Index scores for each Mosaic Type were appended to every postcode in London in order to create a London wide collision propensity map. Figures 4.11, 4.13 and 4.15 show the casualty propensity for Hampstead, Kilburn and part of Willesden. The index scores for postcodes in Hampstead are lower than average and lower than the areas of West Kilburn and Willesden Green. The latter areas are parts of London with high deprivation compared to Hampstead. By contrast Figures 4.10, 4.12 and 4.14 shows the propensity to be involved in a collision as a driver in the same area. The propensity is lower for involvement in collisions as a driver. Hampstead generally is still below average, and areas of West Kilburn, Willesden Green and Maida Vale have a higher than average propensity to be involved as a casualty rather than a driver indicating likelihood that being involved in a collision as a vulnerable road user such as a pedestrian. The maps show postcode points, of which each postcode is appended to a Mosaic Type. Based on the casualty and driver postcode data, the index scores for those casualty and drivers were appended to all the postcodes and therefore Mosaic Types in London. Therefore the maps shown are evenly distributed and a calculated 'potential' risk, based on the high numbers of people in their Mosaic Type being involved in collisions. The drawback of this approach is that it avoids any possibility of examining within group variance and only focuses on all collisions.

[4] Conjecture

Figure 4.10: *Driver's living in Hampstead propensity of being involved in a collision*

Figure 4.11: *Casualty's living in Hampstead propensity of being involved in a collision*

In the area of West/East Ham in London the pattern is very different. Figures 4.10 and 4.11 show the propensities of both drivers and casualties to be involved in a collision as either a driver or a casualty from this area. There are virtually no postcodes which have an under representation to be involved as a casualty, with only a few postcodes lower than average as a driver in the south west corner of the map. This area has a strong proportion of type D27 'Settled Minorities', D26 'South Asian Industry' and F36 'Metro Multiculture' all of which were identified as having a higher the average likelihood of being involved in a collision. There is a significant number of this type clustered in one area indicating a very overall high propensity to be involved in a collision coming from this area of London.

Figure 4.12: *Driver's living in West Ham propensity to be involved in a collision*

Figure 4.13: *Casualties living in West Ham propensity to be involved in a collision*

In a direct comparison the area of Kensington and Chelsea offer a very different pattern to that of West/East Ham. Figure 4.15 shows the index scores for the likelihood of being involved as a pedestrian in a collision. Nearly all of the postcodes are over represented, with a small proportion of over represented postcodes near North Kensington, Notting Hill and the A40. The same map seen in Figure 4.14 shows the propensities of being involved in a collision as a driver and shows all the postcodes within the map having an under representation of this occurring. This area of London has, similar to Hampstead, high numbers of Group A 'Symbols of Success', particularly type A01 'Global Connections'. These Groups has very low index scores with regards to driver and casualty propensity.

Figure 4.14: *Driver's living in Kensington propensity to be involved in a collision*

Figure 4.15: *Casualties living in Kensington propensity to be involved in a collision*

From the spatial analysis there are clear and distinct areas in London which have residential areas where residents are more likely to be involved in a collision as either a driver or a casualty. From the results we can see that there is a higher proportion of residential postcodes whereby the resident is more likely to be involved in a collision as casualty rather than a driver. This is predominantly because of the mobility patterns within London whereby people are predominantly involved in pedestrian or cycling activities rather than driving themselves. There is a tendency for Mosaic Types which experience higher deprivation and financial constraints to be more likely to be involved in a collision as a casualty. Transport for London in conjunction with the Government is keen to address the link between social exclusion and road safety through partnership working schemes, particularly involving local boroughs.

4.6.4 Regional effects – geodemographics and road collision propensities

An implicit assumption behind this analysis is that all the characteristics of Mosaic Types are constant across the UK (Webber 2005). However it is clear from the descriptions of the Types that they are not evenly spatially distributed. For example, Type J56 'Tourist Attendants' have a strong propensity to be located in coastal areas such as Whitby, Torbay and Newquay. The percentage of these types of people is almost negligible in London. In comparison, Type A02 'Cultural Leadership' has 56% of its postcodes within London, in areas such as Muswell Hill, Finchley and Wimbledon. Therefore although the spatial location of these groups is uneven, the characteristics of the types are assumed to be equal, whether Type A02 is in Richmond, Aston or Sheffield. Actually equating the indexes across a larger area and at a larger scale in terms of population measured should allow the user to have more confidence in the index values. Therefore the larger the sample the more robust it should be. However things will happen above and beyond what Mosaic can classify and there are no doubts that regional effects exist (Fotheringham *et al* 2000). The question is therefore, how much do these regional effects matter? The answer would be that they probably do matter for some areas and some variables, while for others it may be that discriminating by Mosaic is better than by borough, age, sex, ethnicity or any other socio-economic/geographical variable.

There is a continuous drive by the government for continued community based intelligence for public services (Longley 2005). This has been most prominent in the field of both crime and health (Ashby *et al* 2005, Levy *et al* 2005). These studies provide analysis of in-depth community similarities and differences in terms of say health targeting or local policing needs. However there is scope for research into potential geographical and regional effects above and beyond those accounted for by national geodemographic typologies (Ashby *et al* 2005). In this section, the differences are discussed between using a London wide base population or a borough base population when equating index scores for a particular borough. In theory by using the borough base population and associated Mosaic counts for each postcode the index scores should be more accurate. In examples in the next section the index scores and spatial differences are discussed. Three boroughs in London have been selected to highlight the possible regional disparities. Camden, Croydon and Hillingdon have been chosen as they represent boroughs in different areas of London: one is located in Central London; Croydon is located on the outskirts and has a large town centre; and Hillingdon lies on the edge of Outer London (see Figure 4.16).

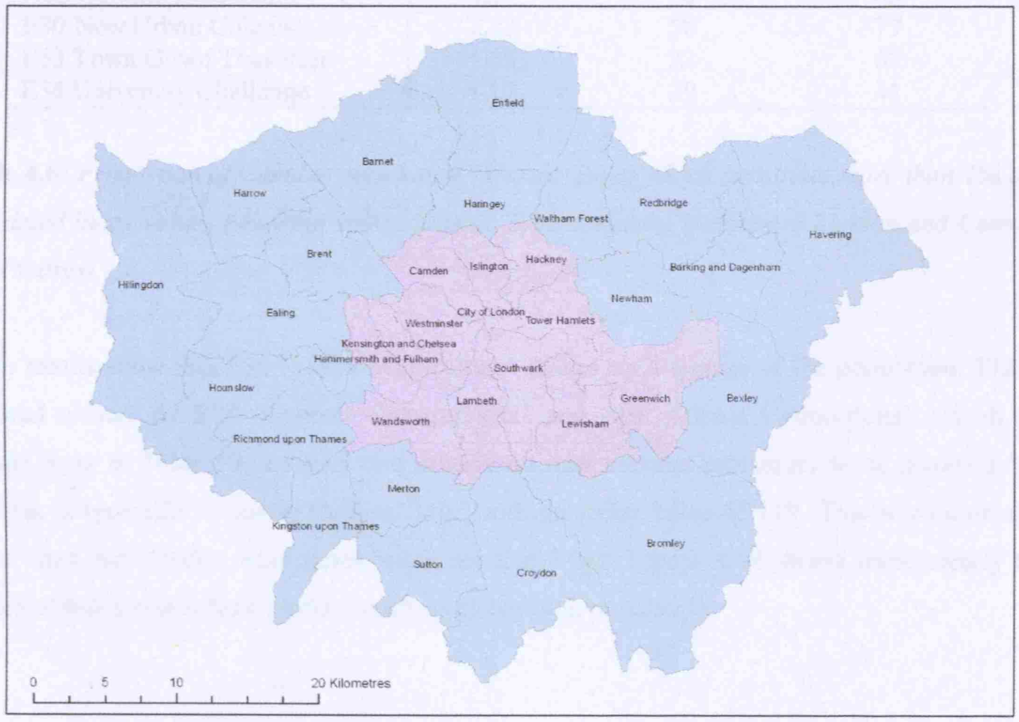


Figure 4.16: *Map to show the location of Camden, Croydon and Hillingdon*

The method mirrors that outlined in Section 4.6.3. However, a fundamental difference between the two is that the base population here is for the borough, rather than the whole of London. This means that the index scores for the casualty propensities will be based on the individual borough's population structure, since this is taken to be the level at which interventions might be formulated. The following section discusses the differences between the London and borough index scores and maps the associated index scores.

4.6.4.1 Camden

This borough has a polarised population in terms of the number of Mosaic Types; there is a large percentage of the population which falls into a small number of types. Table 4.6 displays the Mosaic Types which have a proportion of the borough population over 1% and the associated casualty Mosaic codes and London index scores for comparison.

Mosaic Type	Camden %	Camden Index	London Index
F36 Metro Multiculture	25.20	97	101
E28 Counter Cultural Mix	24.25	119	91
A01 Global Connections	22.83	93	42
E29 City Adventurers	12.09	99	68
A02 Cultural Leadership	4.87	80	63
E30 New Urban Colonists	2.23	58	77
E33 Town Gown Transition	1.53	22	67
E34 University Challenge	1.17	29	41

Table 4.6: *Proportion of Camden population (Mosaic Types which constitute more than 1% and associated index values (showing within Mosaic Type variation if compare London and Camden index scores)*

These results show that F36 'Metro Multiculture' makes up a quarter of the population. This is followed closely by E28 'Counter Cultural Mix' and A01 'Global Connections'. Of all the Mosaic types in Table 10, the only one to have an over average propensity to be involved in a collision is type E28 'Counter Cultural Mix' with an index value of 119. This is considerably higher than the London wide index value for this Type. Figure 4.18 shows more clearly the pattern of index scores for Camden based on its borough population.

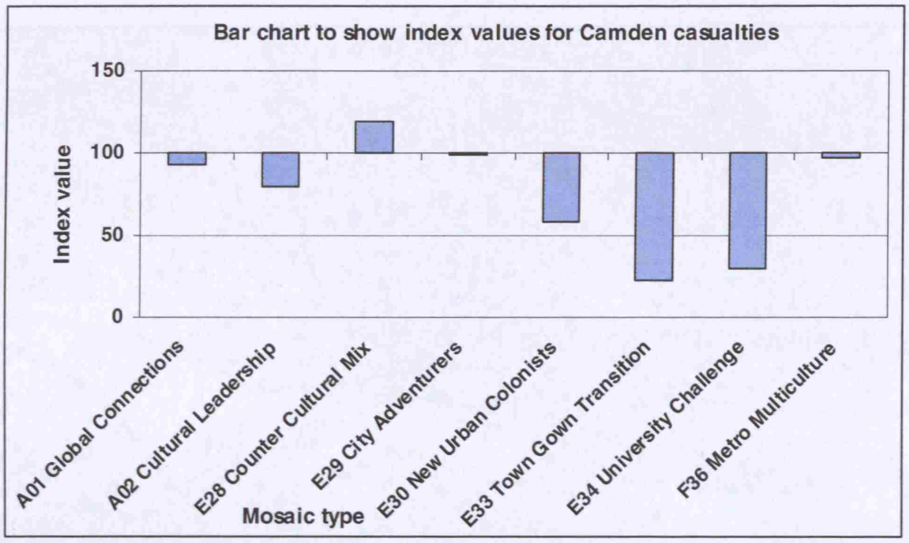


Figure 4.17: *Bar chart of Camden casualty index values*

This bar chart (Figure 4.17) shows that relatively few Mosaic Types have an above average propensity to be involved in a collision. However nearly a quarter of the population has an index value of 119 which can be seen in Figure 4.18. E28 'Counter Cultural Mix' is characterised by young professionals in rented flats, perhaps ethnic minorities sharing old houses and poor tenants in council flats. Most of these people will live in flat and have little access to a car of their own, indicating a strong propensity for this group to be involved in collisions as casualties. Due to the nature of their dwellings, these people will often meet people in cinemas, bars and restaurants indicating a high mobility around the local area[5].

[5] Conjecture

Figure 4.18: *Camden residents and their regional effects to be involved in a collision using a Camden base population*

Figure 4.19*: Camden residents and their regional effects to be involved in a collision using a London base population*

The difference between the two maps in Figures 4.18 and 4.19 is the increased propensities when using the London population as a base. The patterns are generally similar; however using the Camden base population the index scores generally tend to be slightly lower.

4.6.4.2 Croydon

Croydon has a very different population to that of Camden. A fifth of its population is made up with type D27 'Settled Minorities'. Table 4.7 lists the proportions of the population in each Type and the associated index scores.

Mosaic Type	Croydon %	Croydon Index	London Index
D27 Settled Minorities	20.80	104	113
C19 Original Suburbs	11.11	82	96
C20 Asian Enterprise	5.63	109	105
D21 Respectable Rows	5.35	114	119
H46 White Van Culture	5.28	132	135
A03 Corporate Chieftains	4.87	83	86

E28 Counter Cultural Mix	4.27	61	91
A05 Provincial Privilege	3.99	62	82
E30 New Urban Colonists	3.48	63	77
E32 Dinky Developments	3.31	111	126
H47 New Town Materialism	2.51	128	169
G41 Families on Benefits	2.46	127	142
A06 High Technologists	2.34	101	163
E29 City Adventurers	2.31	81	68
F36 Metro Multiculture	2.31	130	101
J52 Childfree Serenity	2.12	75	81
C15 Close to Retirement	2.07	102	146
C18 Sprawling Subtopia	1.92	123	111
A02 Cultural Leadership	1.76	63	63
B11 Families Making Good	1.75	142	185
B12 Middle Rung Families	1.54	176	151
A04 Golden Empty Nesters	1.45	89	124
A07 Semi-Rural Seclusion	1.12	125	282

Table 4.7: *Mosaic index values for casualties from Croydon*

The first point to notice in Table 4.7 is the large proportion of Mosaic Types which each individually account for over 1% of the population, compared to the Camden population. Type D27 'Settled Minorities' are characterised by second generation ethnic minorities moving out of the inner city looking for more attractive accommodation. These people are likely to catch buses into Central London to go shopping and therefore to have a high pedestrian mobility, to and from bus stops and using local shops. These people want to be seen to be driving in prestigious vehicles, however it is often old BMW's which litter the driveways. This type has a slightly lower index score than the London average. Three other significant types which have a large proportion of the population and an above average index score are C20 'Asian Enterprise', D21 'Respectable Rows' and H46 'White Van Culture'. Type D21 'Respectable Rows' is characterised by childless couples who are young and in the initial stages of cohabitation. This type usually has access to a car and sometimes two; often it is a company car, but if it is privately owned it would be an expensive model. In this borough there seems to be strong differences between the mobility of the different population Types, for example D27 'Settled Minorities' are likely to be involved as a pedestrian, whereas D21 'Respectable Rows' are much more likely given their car ownership and use of the local area be involved as driver casualties.

Figure 4.20 below shows the index values for the Mosaic Types in Croydon which make up 3% or more of the population. This graph shows a strong propensity for H46 'White Van Drivers' as well as E32 'Dinky Developments' to be involved in a collision as a casualty. The latter type would predominantly use the bus or trains to travel.

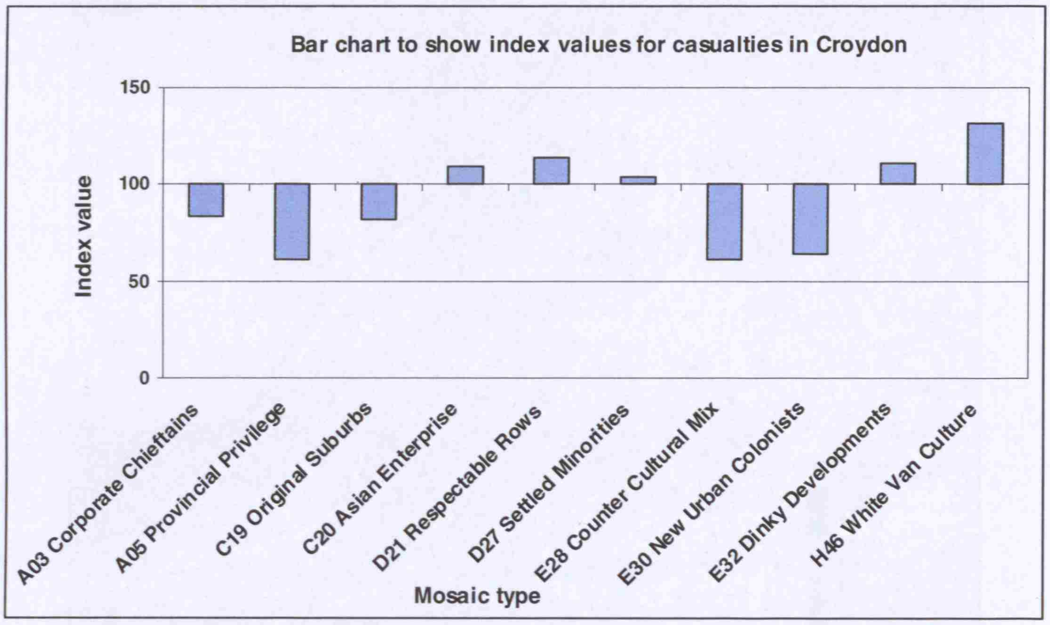


Figure 4.20: *Bar chart to show the casualty propensities by Mosaic Type in Croydon*

Figure 4.21: *Croydon residents and their regional effects to be involved in a collision using a Croydon base population*

Figure 4.22: *Croydon residents and their regional effects to be involved in a collisions using a London base population*

The areas of low collision likelihood are more prominent on the Croydon base population map than the London base population map (Figures 4.21 and 4.22).

4.6.4.3 Hillingdon

Hillingdon shares similar characteristics with Croydon with regard to the high number of Mosaic Types which occur within the borough. It is apparent from Table 4.8 that the make up of the Borough with regards to Mosaic Types is focused more on suburban residents, highlighted by the high proportion of type C20 'Asian Enterprise', H46 'White Van Culture', C19 'Original Suburbs' and C18 'Sprawling Subtopia'.

Mosaic Type	Hillingdon %	Hillingdon Index	London Index
C20 Asian Enterprise	12.59	96	105
H46 White Van Culture	10.23	101	135

C19 Original Suburbs	8.68	97	96
C18 Sprawling Subtopia	8.50	107	111
D21 Respectable Rows	7.42	91	119
A05 Provincial Privilege	4.93	72	82
D27 Settled Minorities	4.46	122	113
E32 Dinky Developments	4.43	103	126
B12 Middle Rung Families	4.30	96	151
J52 Childfree Serenity	3.02	94	81
B11 Families Making Good	2.77	105	185
C15 Close to Retirement	2.47	110	146
B13 Burdened Optimists	2.34	128	213
A03 Corporate Chieftains	2.23	66	86
H47 New Town Materialism	1.97	131	169
A04 Golden Empty Nesters	1.88	90	124
G41 Families on Benefits	1.83	126	142
A02 Cultural Leadership	1.72	86	63
E30 New Urban Colonists	1.26	109	77
A06 High Technologists	1.17	85	163

Table 4.8: *Table to show the population of Hillingdon by Mosaic Type and associated casualty index scores.*

Of these Mosaic Types very few have higher than average propensities to be involved in a collision apart from C18 'Sprawling Subtopia' which has a slightly higher than average propensity with an index score of 107. Type D27 'Settled Minorities' has one of the highest index values of 122. Although this type made up a large proportion of Croydon's population it did not have a higher than average likelihood to be involved in a collision, unlike Hillingdon. There are some prominent differences between the Hillingdon index scores and the London index scores. For example type H46 'White Van Culture' has a London wide index score of 135 compared to a considerably lower Hillingdon score of 101. Compared to Croydon the difference between the two base population scores are very different. Croydon's index scores are rather uniform, whereas those of Hillingdon show greater variability. As a general pattern the London base population index scores for Hillingdon tend to be higher than when using the Hillingdon population as the base.

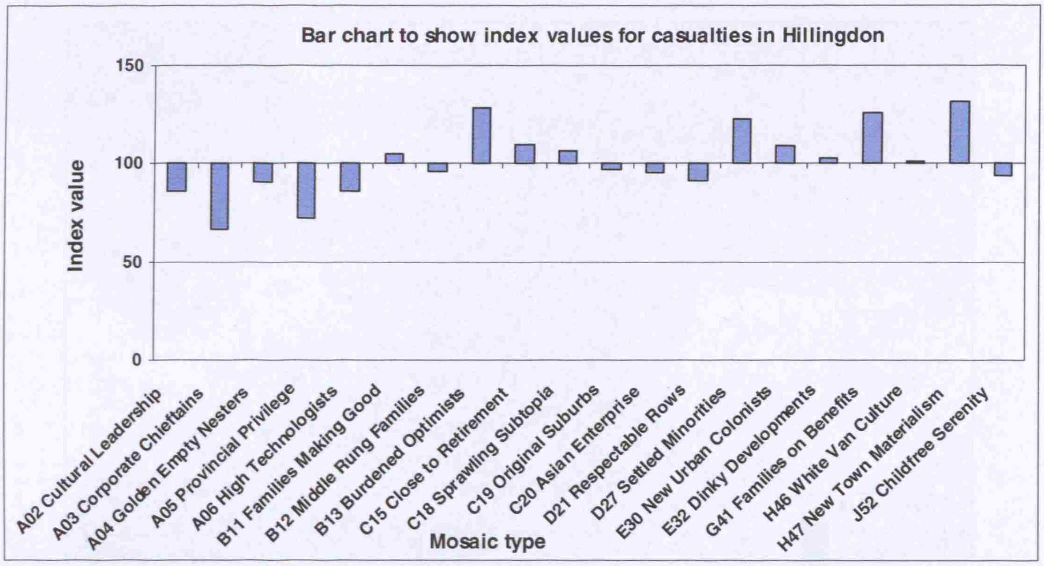


Figure 4.23: *Bar chart to show the casualty propensities by Mosaic Type in Hillingdon*

Figure 4.23 highlights the low propensities of being involved in a collision of Group A ('Symbols of Success') and Group B ('Happy Families'). There is an increase in propensity for Group C ('Suburban Comfort') which is expected with regard to the nature of the location of Hillingdon. Figures 4.24 and 4.25 show the spatial differences of the index values using the two different base populations. More similar to the differences in the Camden maps, these two maps of Hillingdon show that using the London base population index scores creates a more intense with higher index scores in general.

Figure 4.24: *Hillingdon residents and their regional effects to be involved in a collusion using a Hillingdon base population*

Figure 4.25: *Hillingdon residents and their regional effects to be involved in a collision using a London base population*

Overall the differences in the index scores for the three boroughs are small: however the results do show that there are differences in detail.

4.7 Conclusion

Arguably geodemographics offers a worthwhile and suitable solution to a complex problem. Socio economic areal solutions to profile road collision victims have been slow in evolving. There has been considerable criticism with regard to the application of road collision data to geodemographics, with regard to the notions of time, mobility and static versus dynamic risk. However, measuring a person's whole risk of being involved in a collision is difficult. There are many factors to consider such as mobility, or how the person interacts with the road (i.e. as a driver, pedestrian or cyclist). This collision risk can change over time, and often does. The limited analysis of casualties postcode data derives from the fact that is does not need to be collected for the STATS19 data records. It would be useful for future research if this collection procedure could be enhanced.

In the work by Abdalla *et al* (1997), the unit of analysis used is not the postcode of the casualty's address but the 1991 Census Output Area (OA), in order to analyse the deprivation of an OA in relation to the casualty address patterns. Geodemographics is more sophisticated, and as well as using many census variables it uses credit histories and qualitative research.

It has often been asked whether or not the collision database can be disaggregated, in order, for example, to ascertain the number of cyclists and work out index scores for each Mosaic Type based on cyclist collisions. However the limitation with this and the reason why it has not been attempted here is that it would give an inaccurate measure of risk because the actual number of people who own or even ride their bicycles in each Mosaic Type is not known which means care needs to be taken when discussing results. Driver propensities are marked by those people who are living in Outer London and therefore have a greater use of cars for personal and business use.

This chapter has sought to understand in more detail the characteristics of the casualties of road collisions based on their postcodes and applied the results to a London wide spatial analysis in order to ascertain the overall collision propensity patterns for London as a whole city. These results show clear patterns of 'residential hotspots' whereby the residents exhibit higher than average propensities of risk. Perhaps this highlights the main drawback of using geodemographics which is that it is static in time. It is difficult to conclude the Mosaic Types and their changing propensities across time because we do not have information with regards to where they are on a small scale temporal basis (e.g. A day). The conclusions of this chapter reveal neighbourhood patterns of collision propensity. It should be made clear that not everyone in these groups will experience the same collision propensity; however the results here give a good impression of the level of risk involved. Another point to consider is trying to interpret or even break down the index score into a real life situation of potential risk for road users. This would be an area of research to elaborate in the future. The next section concentrates on the other spatial location associated with collisions, and that is the collision location itself. It aims to explore the nature of collision hotspots and appropriate methods for identifying these areas of collision risk.

CHAPTER 5

DETERMINING ROAD COLLISION HOTSPOTS IN LONDON: A SURFACE BASED APPROACH

5.1 Chapter overview

Throughout this thesis, emphasis has been placed upon identifying the dominant variables that contribute to a collision, specifically when the incidence of collisions appears to be spatially dependant (i.e. they share a similar space and 'closeness' or occur in what is known as a 'blackspot' 'black zone' or 'hotspot'). One of the strengths of this approach is that in this thesis I am measuring proximity not only in a physical/geographic space but also in a social space of the Mosaic classification. Road collision hotspot analysis has traditionally centred on road segments or specific junctions (Thomas 1996) while area wide hotspots and the spread of risk which is produced from a collision is somewhat neglected. This chapter is based on the assumption that road collisions occurring in a similar area, not just taking into account the road network or junctions, are spatially dependant because of the increased density of collisions in a specific area. This dependence is argued to be the result of a shared common cause(s) between the collisions, albeit of varying intensity. Causes could include similar environmental features, collisions occurring at a similar time or involving people who share similar characteristics.

This chapter describes a two tiered research methodology that aims to develop a meaningful hotspot analysis based on collision density over space and vision. The first section of the chapter explores the applicability of kernel density estimation as an appropriate method for identifying collision hotspots. The second section of the chapter outlines the surface results from the kernel density estimation techniques.

SECTION A

5.2 Road collision surface analysis

5.2.1 Introduction – Towards a surface based approach

Traditionally, mapping using data point features has been carried out through representing populations as individual point objects using symbols, or colours to differentiate between the data values held by the point. It is not often within socio economics that precisely referenced points are available for reasons of confidentiality. More recently there has been a move towards the use of surfaces to model the likely distribution of the original points in a more useful and visually understandable format (following Haggett *et al* 1977; Martin 1996; Martin 2000). The conclusions from these papers which outline the advantages of using a surface based approach indicate that surface representations can usefully summarise the distributions of point referenced events when point locations are known.

The aim of this section is to outline and demonstrate the importance of using density measures rather than count measures of collision data. A density is the amount of something per unit area, expressed relative to some meaningful base category, for example the total population per unit area within an output area or borough. The density is defined as all points provided one specifies what neighbourhood or region around a point should be used to summarise the data. For example, the population density of an area would be based on the area (in metres for example) of the surface of the boundary data, together with the count data (for example, the population). To create comparable densities at different locations, a comparable distance weighting function around each point is chosen.

The conceptual basis to the first section of this chapter lies in the premise of the spread of risk that a collision generates over an area. In many circumstances, it has been acceptable to define a cluster of collision points in an area or on a line, but the risk of a collision occurring again will likely spread beyond the boundaries of the historical collision cluster. If the cluster of collisions is small and spiky the risk surrounding the cluster will be smaller, however if the collision cluster is quite flat but covers a wider area then the risk surrounding the cluster will be larger but less intense than the cluster with the collisions closer together (Chainey and Ratcliffe 2005).

Although there has been some use of kernel density estimation (Sabel *et al* 2005) for analysing road collision point data there has been limited explanation of the method in the road safety literature. For example the optimum bandwidth discussion for collision data has been neglected. Its use in recent years in social science has focused on representing the density or volume of crimes distributed across the study area (Chainey and Ratcliffe 2005, van Eck *et al* 2006). Chainey and Ratcliffe (2005) suggest that what is still missing from the generic kernel density estimation is a method that can statistically define those areas which are hotspots. Determining hotspots in this manner is subjective: however using clearly defined and meaningful parameters it can produce useful and interesting results which are shown throughout the next two chapters. The practical options surrounding the method mean that the choices made in this chapter have taken into consideration results of other studies and applications – for example, the town centres study of Lloyd (2005). The town centres project utilised a subjective method to determine the optimum bandwidth for use in town centre definition, namely the maximum distance that a person was deemed willing to walk in order to reach the centre of a town. This subjective decision making process presents greater challenges when applied to measuring collision density surfaces, and the methodology will be discussed in the following sections.

5.2.2 Previous literature: identifying road collision hotspots, methods and techniques used

The collision literature provides no universally accepted definition of a road collision 'hotspot'. Hauer (1996) describes how some researchers rank locations according to collision rate (this is usually collisions per vehicle kilometre), while other researchers use collision frequencies (collision per road kilometre). Another dimension for contestation is that rank may be determined by the magnitude of either rate or frequency or by the amount by which the rate or frequency exceeds what is 'normal' at a specified range of sites.

Sayed and Abdelwahab (1995) suggest that the number of collisions is affected by three factors, which can be identified below:

- The road environment
- The condition of the vehicle using the road network
- The concentration, skills and physical well being of the road user

The models outlined in the next few sections generally explain collisions as being a consequence of driver behaviour that is not correctly matched to the demands of the road environment or to the vehicle itself. The demands of the road vary because of many different factors, including traffic flow, time of day, geometric features, and road user characteristics. What usually occurs is that the road user (including both the driver and perhaps say, the cyclist, pedestrian etc) will adapt their behaviour to suit the demands of the road (Geurts and Wets 2003). Hotspots have been argued by Geurts and Wets (2003) to be areas or points that demand peak driving performance of road users.

5.2.3 Statistical models

Ever since the late 1970s, basic statistical models have been applied to road traffic collision analysis. For example, Foldvary (1979) and Jovanis & Delleur (1983) each used simple models using mean and variance, focusing on the variations in collision rates. However, the main drawback of these models is that they are unable to incorporate the effect of risk factors on the collision involvement. The statistical history of collision modelling follows a basic pattern of events, which as mentioned begins with exploring collision rates using simple mean and variance models and in the mid 1980s moves towards using multiple regression techniques to explore road collision patterns. For example Oppe (1979) and Ceder & Livenh (1982) adopt multiple linear regression models using a dependant variable such as the collision frequency or rate and independent variables describing the characteristics of the collision or traffic environment -- such as traffic flow, speed, type of road and severity. These models have a number of drawbacks including the assumption that collisions are assumed to be normally distributed – this is rarely appropriate when describing the random, infrequent and discrete nature of road collision events.

The next step to accommodate this statistical drawback was to use Poisson log linear regression to account for the randomness of collisions in time and space. Saccomanno & Buyco (1988) and Blower *et al* (1993) both used Poisson log linear models to explain the variations in collision rates. However this particular model failed to account for the extra Poisson variation (the value of the variation could exceed the value of the mean) in the collision count data. Many authors such as Maycock & Hall (1984), Hauer and Persaud (1987), Persaud (1990), Miaou (1994), Shankar at al (1995), Kulmula (1995), Maher and Summersgill (1996), Hauer (1997), Tunaru (1999) and Abdel-Aty & Radwan (2000) have used negative binominal regression models. In all of these models only the reported number of collisions in the observed time period is used and locational characteristics are therefore modelled as constant within a given time period. However, characteristics such as traffic flow do change over time and therefore it is important to be aware of this: for example in the methodology described for this study, the dataset is divided into different levels of temporal disaggregation, such as hour of day, day of week, week of year, month and year. This means that the changes in the hotspots can be monitored at different temporal levels in order to assess the nature of the changes and how they might be managed. It is also worth noting that the majority of these collision models are specific to certain road network characteristics such as intersections, traffic lights, roundabouts, or specific vehicle types such as trucks or motorcycles.

5.2.4 Selecting sites for further investigation

The importance of ranking hotspots has not been neglected in the literature. It is a fundamental process of road safety treatment. A significant part of this chapter focuses upon the importance of ranking areas of collision and upon the criteria that are used to rank them. Further investigation can mean anything from more in-depth statistical analysis of the collisions occurring at a particular site to alterations of 'road furniture' and environmental characteristics of the sites such as traffic lights, pedestrian crossings or road signs. One of the aspects of this thesis is the importance of identifying the characteristics of people who have had similar collisions in terms of education and awareness. The tasks involved in identifying and treating a hotspot can be subdivided into three distinct phases (Visitsen 2002):

1. Targeting hotspots on the road network
2. Prioritising the hotspots to treat with whatever measures are deemed appropriate
3. Before and after studies of the effect of the treatment

Although the treatment of hotspots through engineering methods is not being studied in this thesis it is important to be aware of the basic approaches which are commonly used. The Australia Bureau of Transport and Regional Economics outlines a four stage approach which aims to reduce crashes:

1. Single sites or hotspots treating specific sites or short sections of road

2. Route countermeasures
3. Area wide action
4. Mass action

Ranking of hotspots usually entails choosing locations which exceed a chosen threshold value, and constitutes an important part of this thesis. This method has its drawbacks as it is very sensitive to the random variation in collision counts. In the past there have been many different methods for identifying the threshold number beyond which to rank a site or stretch of road. For example McGuigan (1981) suggested ranking sites according to their potential for collision reduction, which is the difference between the reported number of collisions and the expected number at locations with similar characteristics. In a subsequent study Persaud (1999) used empirical Bayesian estimation. The task of targeting hotspots can be viewed by some as a ranking and selection problem, although Gupta & Hsu (1980) have also introduced the concept of probability of correct selection (PCS). Here, a subset of a group of locations is targeted as hotspots, if the probability of selecting the site with the largest expected number of collisions (the 'worst' location) is above a chosen threshold value.

Interestingly Van den Bossche (2002) completed an investigation into whether ranking alone can provide enough evidence for the selection of dangerous sites. The authors use Bayesian hierarchical modelling techniques which are used to identify and rank hazardous sites (intersections for bicycles in Belgium). The authors conclude that ranking hazardous sites is an interesting means to gain insight into dangerous locations but there no such thing as a 'correct' ranking system.

In an ideal situation for the treatment of hotspots, there would be unlimited funds in order to improve the safety of the hazardous locations. However, there is of course no such thing as an ideal situation. Implementing safety measures is costly and there is usually restricted funding with which to treat hotspots, which therefore means that not all of the sites can be fully treated. This leads to a challenge of prioritising sites for treatment (Geurts & Wets 2003). This in turn involves cost benefit analysis in terms of putting a 'price' on a collision, which can be a difficult task, and also knowing whether the reduction of collision at a site was specifically because of the implemented reduction measure or other factors (such as the temporal ebb and flow of collision hotspots).

5.2.5 Profiling the hotspots

Central to the underlying conceptualisation of this thesis is the quest to obtain a deeper understanding of the processes and risk patterns which are apparent at ranked hotspots. Profiling hotspots can lead to a more beneficial outcome for reduction, and is therefore extremely important to policy makers and analysts alike. Road collisions are often linked by a number of common features that are manifestations of spatial dependence, temporal dependence and shared characteristics of the people involved in collisions. It is often the case that of these different features only one is examined at any one time. For example the severity of the collision (fatal, serious, and slight) might be investigated, or the collision contributory factors which occur at the scene of the collision and the possibility of the collision category such as collisions involving pedestrians or the elderly.

5.2.5.1 Profiling according to severity

Modelling collisions by severity has often proved a more obvious choice when measuring risk levels of hazardous sites because the surface portrays that the larger the number of fatal collisions at a specific location the higher the risk and therefore importance of that site. A number of models have been designed to integrate severity (see, Gimotty & Chirachavala 1982, Latimer 1992, Nassar *et al* 1994 and Wood & Simms 2002). Various models such as binary logit regression are used: thus Gimotty & Chirachavala (1982) investigated the relationship between occupant injury severity and collision conditions, using the model to differentiate between severe and non severe injuries. Nassar (1994) develops an integrated Accident Risk Model (ARM) for policy makers using risk factors affecting collision involvement on road sections and injury severity of occupants involved in the collisions. This model is based on negative binominal regression.

5.2.6 Risk factors

Risk factors are used to explain collision involvement and collision severity. According to Nassar (1994) risk factors in collision models play two roles:

1. Improving the overall model fit and reducing the amount of unexplained variation
2. Providing a means for evaluating the effectiveness of alternative safety measures

Nassar (1994) also elaborates on the different risk factors associated with collision involvement:

- Cause of the collision, for example vehicle manoeuvre, driver action

- Traffic conditions, volume, speed
- Environmental conditions, such as light, weather, road surface
- Human conditions, driver age, mood, driving experience, sex, seat belt use
- Vehicle conditions, type of vehicle, engine size, brake conditions

In terms of risk factors influencing severity, the relationship is strong and the studies mentioned have determined strong relationships between risk factors such as speed, seat belt use and age to be very important in terms of explaining collision severity.

5.2.7 Identification of hotspots

Identification of hotspots in the literature has been seen as arising from the awareness of the spatial interaction existing between collision locations. The existence of these hotspots reveals concentrations and hence suggests a degree of spatial dependence between the point locations. These spatial concentrations could arise because of one or several causes of collision.

Thomas (1996) argue that the most appropriate level of spatial aggregation for road collision analysis is the road section in terms of a pre determined segment of road; however in most studies this length is not controlled or even justified. There is no clear indication in the literature as to what would be the most meaningful length for a 'dangerous' section of road: however, this does call into question whether there should be an optimal length of road section to analyse as hotspots and black 'segments' would vary in size and length depending on the characteristics of the individual hotspot, therefore suggesting that to determine an overall length to be used for each black 'segment' would potentially miss out individual collisions that may nevertheless share spatial dependence with other collisions in the hotspot.

Flahaut (2002) uses the concept of the hotspot to tackle the problem of length and location of road sections. The motivation for this research is purely to try to define the location and the length of what Flahaut (2002) terms 'black zones' rather than the factors which could explain the occurrence of the collisions. Two methods are compared, namely the calculation of local spatial autocorrelation measures and kernel density estimation. Both methods graduate levels of local danger and generate smoothing of the empirical process, and although each of the methods starts from different conceptual backgrounds both provide quite similar results under a specific choice of parameters.

5.2.7 Conclusions

A range of statistical models have been outlined and discussed in this review. In the literature there is no universally accepted definition of what constitutes as a 'hotspot. Hotspots are usually assessed in an ad hoc fashion depending on the policies of the administrative area and issues of locational jurisdiction. Locations are usually classified as hotspots after an assessment of the level of risk and the likelihood of a crash occurring at any given site. Researchers have in the past proposed several alternative methods for targeting and ranking hotspots: however there have been no solid conclusions as to the 'correct' ranking or identification method. Funding plays a large role in the scope to identify hotspots and the resources to employ safety measures. This section has highlighted the ongoing need to explore different methods and outcomes for collision hotspot identification, analysis and targeting as there is no optimal solution to this important health problem.

5.3 Kernel density estimation

Kernel density estimation is an interpolation technique, which is a method for generalising incident locations to an entire area. In short, whereas spatial distribution and hotspot techniques provide comprehensive statistical summaries for the data incidents alone, interpolation techniques generalise the collisions over the study region. There are many interpolation techniques such as Kriging, and local regression models. Kernel density estimation is an interpolation method which is appropriate for individual point locations (Silverman 1986; Bailey and Gatrell 1995). Density is a measure of the quantity of something per unit of area, such as the number of collisions per square mile or the number of people per square mile.

By interpreting the collision point data in the form of a density surface, a number of decisions have to be made in order to facilitate appropriate and robust surfaces and ultimately for this research the results. The literature is characterised by many similar academic studies into road collision patterns processes using the original data points of the collision locations and representing this population as the original points using symbols and possibly varying colours for the types of severity of type of vehicle involved in the collision. These are known as dot density maps. However what these maps fail to show is the spread of risk that a collision generates. There has been limited work which has been focused on the density maps in terms of bandwidth choice and kernel values. This section endeavours to explore this with reference to other studies which have used and manipulated density surfaces and bandwidths.

The use of surfaces has been particularly prolific in crime pattern analysis (Radcliffe 2000, Levine 2005). Within this context it has been useful for crime analysis practitioners to apply surface interpolation to point crime data in order to determine hotspots over a given spatial area (Chainey and Ratcliffe 2005). The advantages of using kernel density estimation include the ease with which it can be implemented using standard GIS software routines, its smoothness and relatively fast convergence rate. With this in mind, this methodology has been adopted for this section in order to determine the spatial density of collisions within London. The advantages of these surface representations particularly of road collisions are that they can provide a more realistic continuous model of collision hotspot patterns reflecting the changes in density which are often difficult to represent using geographically constrained boundary based models such as the transport network or census tracts. These advantages mentioned have motivated the choice in methodology for the collision hotspot methodology, and for the aim of creating a taxonomy using a surface rather than point pattern representation has been chosen because of the homogenous size of the spatial unit.

Over the years there have been a number of spatial tools developed which help in the understanding of the changing geographies of point patterns. The most promising of these tools is kernel density estimation (Sabel *et al* 2000). There are many advantages to the use of kernel density estimation (KDE) as opposed to statistical hotspot and clustering techniques such as K-means. The main advantage for this particular method lies in determining the spread of risk of a collision. This spread of risk, can be defined as the area around a defined cluster in which there is a higher likelihood for a collision to occur based on dependence. This degree of risk would not be measured using the clustering techniques. Secondly by using this density method, an arbitrary spatial unit of analysis can be defined and be homogenous for the whole area which makes comparison and ultimately a taxonomy possible.

Kernel density estimation involves placing a symmetrical surface over each point and then evaluating the distance from the point to a reference location based on a mathematical function and then summing the value for all the surfaces for that reference location. This procedure is repeated for successive points. This therefore allows us to place a kernel over each observation, and summing these individual kernels gives us the density estimate for the distribution of collision points (Fotheringham *et al* 2000).

The concept of this method originated in the 1950s as an alternative method for the density of a histogram. This concept was applied to univariate data but used in a geographical context it is also applicable to multivariate data appreciating the spatial distribution and intensity of the points. Its application to collision analysis will be based on using the x, y co-ordinates for the location and obtaining the density from the count data. The KDE equation can be written as (Fotheringham *et al* 2000):

$$\hat{f}(x, y) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{d_i}{h}\right)$$

Equation 5.1

where $f\ (x, y)$ is the density estimate at the location (x, y); n is the number of observations, h is the bandwidth or kernel size, K is the kernel function, and d_i is the distance between the location (x, y) and the location of the *ith* observation. The effect of placing these humps or kernels over the points is to create a smooth and continuous surface. When computing this method, there are many decisions which need to be made regarding the kernel shape, bandwidth and cell size. These parameters will be discussed in the next section.

Using the kernel density function allows a surface to be created which will visualise the locations of area (or cells) with high (and low) numbers of collisions. Using a spatial density measure for this purpose is more accurate than using count data across space. This is important, as this methodology seeks to determine risk levels at the collision points but also around the points. The count data are used in the initial stages to establish the density which can reflect a spread of risk which may or may not occur around the collision. In the context of this particular method this means that the density of each cell can be exported as a part of a statistical dataset that can be used to identify and analyse the exact cells which have a 'higher than average' density. Due to the low numbers produced by kernel density estimates, the scaling factor simply multiples these small numbers by a constant, in this case it was divided by 4 because the number was to four decimal places. This meant I had a column of Z values whereby a small proportion was over 1. Therefore it was decided that the 'average' would be 1 and I selected the grid cells which had a value over 1 to concentrate on. This means a more spatially accurate method can be created to target collision hotspots and the collisions within those hotspots can be used to determine any patterns with regards to environmental, social, engineering and collision information about the site and collisions within the chosen sites.

The method is known as kernel density estimation (KDE) because around each point at which the indicator is observed a circular area (the kernel) of defined bandwidth is created. This takes the value of the indicator at that point spread into it according to some appropriate function. Summing all of these values at all places, including those at which no incidences of the indicator variable were recorded, gives a surface of density estimates. Density can be measured by two methods; simple and kernel. The simple method divides the entire study area to predetermined number of cells and draws a circular neighbourhood around each cell to calculate the individual cell density values, which is the ratio of number of features that fall within the search area to the size of the area. Radius of the circular neighbourhood affects the resulting density map. If the radius is, increased there is a possibility that the circular neighbourhood would include more feature points which results in a smoother density surface.

The kernel density estimation method uses a mathematically complex way to estimate the density compared to the simple method. The kernel method divides the entire study area into predetermined number of cells. Rather than considering a circular neighbourhood around each cell (the simple method), the kernel method draws a circular neighbourhood around each feature point (the collision) and then a mathematical equation is applied that goes from 1 at the position of the feature point to 0 at the neighbourhood boundary. The chosen radius of the circular neighbourhood affects the resulting density map. If the radius is increased, all other things being equal, the kernel becomes flatter. This kernel function is applied to each collision point and individual cell density values is the sum of the overlapping kernel values over that cell divided by the area of the search radius. A smoother looking density surface is created by kernel density calculations than the simple density calculations. ArcMap 8.2 uses a quartic function to perform the kernel density estimation. The Figure 5.1 below shows how this method works

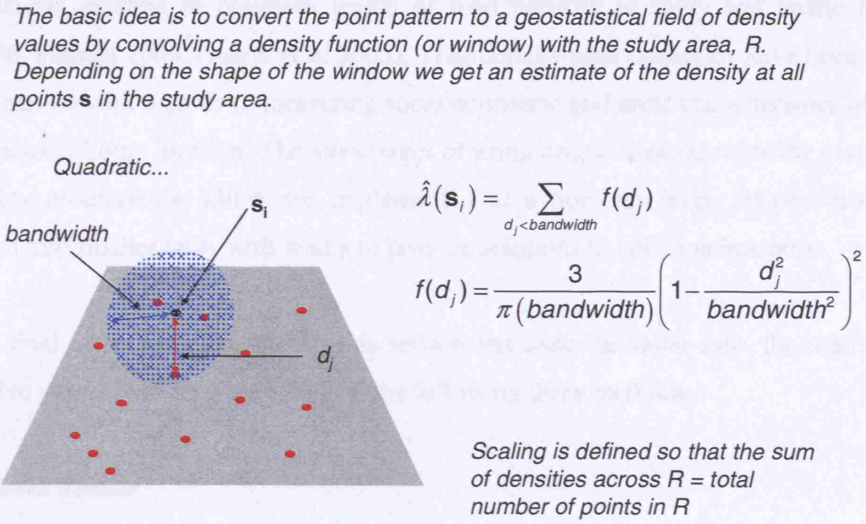


Figure 5.1: *Diagram of how the quadratic kernel density method works and is the basis for the density method used for this study (Source: Cothren 2005)*

5.3.1 Alternatives to kernel density estimation for measuring density of collisions

The alternative methods discussed in this section refer to methods which measure the density of collisions across space. This section of the chapter focuses explicitly on describing and interpreting the results of using a kernel density estimation model to determine the density patterns across London. However, here we see an example of alternatives to kernel density smoothing for road collision data. In order to concentrate on small areas within London and test how geographical constraints affect the locations of road collision hotspots, two boundaries have been chosen. The two alternative methods presented here represent two very different ways to analyse collision density with regards to boundary constraints. Both methods have a boundary constraint in the form of the London road network and Census output area boundaries. The motivation for this section is to compare and evaluate the results of the density output (road network and Census boundaries) compared to each other and the kernel density smoothing method which has no boundary constraints.

Collisions are inherently constrained by the road network and therefore, in some respects it makes sense to analyse collision on the road network. It was outlined in Chapter Three that it is possible to measure spatial autocorrelation between attributes of links on a network (Goodchild 1986, Longley *et al* 2005). A considerable amount of research has been conducted into this issue of collision analysis, relating to optimum length of road network to study and traffic flow (see Thomas 1996, Flahaut 2004, Geurts *et al* 2005). Traditionally road collisions have been linked to census information with regards to measuring socio-economic and areal characteristics of areas in relation to casualty home location. The advantages of using census areas involve the management of road safety programmes which are implemented at a borough level. Often boroughs are disaggregated into smaller units with which to provide solutions to collision hotspots.

As with the final density surface output, this section has used the same data, the road collision dataset for five years. The maps are based on the following three methods:

1. Road network density

Using OS MasterMap Integrated Transport Layer collision point data were counted for each of the 374,000 road segment polylines for London. Using ArcGIS, collisions lying within 5m of a road segment were assigned to a segment. In order to normalise the results the count data were then projected as collisions per 100m.

2. Output area

Collision counts for each Census 2001 Output Area in London were counted using ESRI's ArcGIS. These count data were then normalised by area (in m^2) of the output areas, making a more accurate spatial dataset.

3. Kernel density estimation

In order to see the differences of the hotspot techniques outlined above, a kernel density surface was used in order to make a comparison between the three methods. The kernel estimation used the following parameters: 500m band width and a 200m cell size.

5.4 Road collision surfaces: alternative methods

5.4.1 Results

The maps below show an example of using collision density on a road network and using Census Output Areas. The maps show a crossroads in North London (repeated three times) and show how the three maps (the first one representing the kernel density smoothing method) vary with regards to patterns of collision density. This may suggest that the boundaries in terms of Census Output

Area and the road network restrain or change the location of the hotspot and the hotspots true 'spread of risk'. All three maps (Figures 5.2, 5.3 and 5.4) represent the same crossroads in North London and it is the same scale which makes it suitable for comparison.

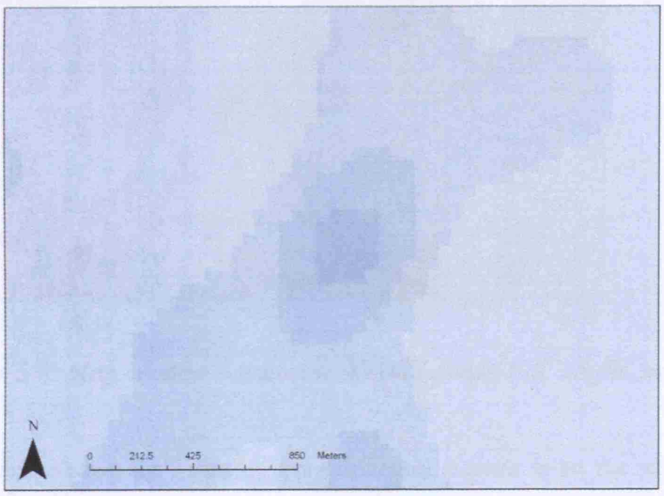


Figure 5.2: *Density map of road traffic collisions at a cross roads in North London (all 'slight' injured road collisions)*

Figure 5.3: *Map to show road segment hotspots at a crossroads in North London (all 'slight' injured road collisions)*

C

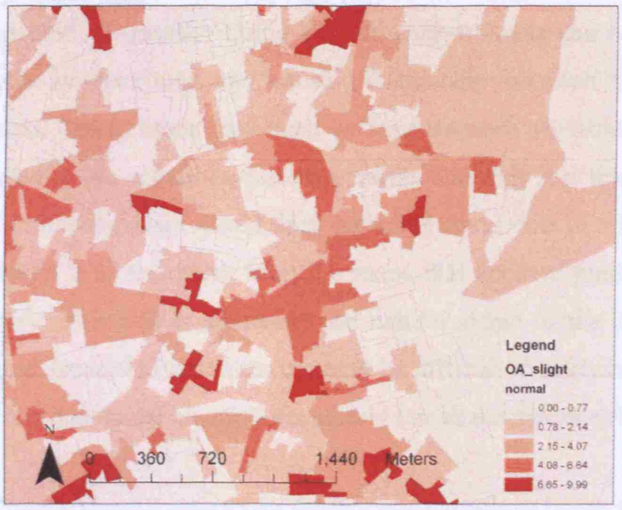


Figure 5.4: *Map to show hotspots in North London (all 'slight' injured road collisions)*

The maps shown here display very different pattern from the same dataset. The density surface (Figure 5.2) clearly indicates four hotspots, whereas the visual depictions on the other two maps are considerably less clear. The road network map (Figure 5.3) shows segments of road which have a high number of collisions per 100m. Although this map highlights a handful of segments it is still unclear as to the location of the hotspot as the collisions were measured per 100m. This method of creating a 'collision rate per unit length' for example per kilometre or hundred metres is a common technique for local authorities. However, there are a number of shortcomings to this method; firstly it disguises the true location of the hotspot along the segment of road because the method uses a rate per unit length. Secondly, although road collisions are spatially restricted to the road network, factors explaining the hotspots may not be directly associated with the road segment or network. For example explanatory factors could be associated with the socio economics of the area, the characteristics of the population that drive in that specific area or the usage of the road (whether it is for heavy goods transportation, major route to shopping complex, or a residential road). Thirdly, a considerable number of collisions occur at junctions, whereby using this method masks these patterns as it is restricted to segments of road.

The final Census Output area map (Figure 5.4) shows inconsistencies with the other two maps. The patterns identify different areas as being subject to high numbers of collisions. This method is clearly more suitable to studying the home location of the casualties and characteristics relating to the area. Both the ITN (Integrated Transport Network) and the output area census maps provide advantages for studying relationships between collisions and data that are available for that specific geography. Using Census Output Areas can result in using many different census variables to determine any causal relationship between socio economic indicators and road collisions. Using other data such as road network variables (e.g.. speed camera location data) appended to the census output area means that they lost their spatial attributes in relation to the collision because they would only represent count data in a larger spatial entity (the Output Area boundary). It is also clear from the maps that because predefined boundaries are used (the OA boundary); this will compromise the natural shape of the hotspot. Similarly, by using the ITN layer and associated attributes it would be difficult to analyse segments of road and append socio-economic data to the segment accurately due to the different levels and shapes of data collection.

This section has described and applied three different spatial methods in order to define hotspots within London. From the results presented, there are advantages and disadvantages of all three methods. Using only a road network for analysis of road collision cluster analysis neglects a vital component of road collision locations, those collisions which occur at junctions. KDE is more suited to locating area clusters including hot 'routes', 'zones' and 'spots'. A network analysis merely concentrates on the route hotspots. By creating a density network, the possible high numbers of collisions occurring at junctions is overlooked and is 'smoothed' and placed into the overall density of that particular route. The importance of junction collisions cannot be underestimated, especially in London. For example in 2003, collisions at or within 20 metres of a junction continued to account for 74.5% of the total number of collisions (Accidents and Casualties on London's roads 2003, Transport for London Annual Report 2003). Using administrative areas for road collision analysis ignores the road network explaining road collision patterns and it also makes the actual collision locations undetectable. Whilst this method gives a good indication of the broad area wide spatial patterns it gives little indication of the local patterns of collision occurrence. Therefore these two methods have been discarded in favour of use a density function within ESRI's ArcGIS.

5.5 Selecting bandwidths and parameters

5.5.1 The selection framework

The choice of bandwidth for this methodology is crucial and will have implications for the results. The bandwidth is the search radius within which intensity values for each point are calculated. Points are weighted, where collisions closer to the kernel centre contribute a higher value to the cell's intensity value of the cell (Ratcliffe 1990). The choice of bandwidth will affect the outcome of the hotspots, for example the larger the bandwidth the larger the hotspots will be. The bandwidth could cover an area the size of a borough or the size of a street, and this in turn will effect the location and size of the hotspots. Some degree of aggregation and smoothing is required if we are to identify urban hotspots which may or may not be restricted to the road network constraints in so far as they may occur at junctions, an area where the road network meets. The choice of bandwidth is important as road safety engineering reaches saturation point in terms of its efficacy in further reducing collisions at certain points on the network, a more local small area approach needs to be addressed.

An important and fundamental proposition of this research concerns the decisions made at the initial stages, namely the kernel density estimation criteria. This study utilises the function in ESRI's Arc products, whereby kernel density estimation (KDE) allows the interpolation of local density values of attributes or point intensities: the statistical basis supporting this tool is described by Silverman (1986). With the aim of investigating the changing shape, form and internal geography of town centres across the nation, kernel density estimation was used in the Town Centres Project as a tool to help explain changing retail patterns and trends (Lloyd *et al* 2004). Town centres, in this context share a similar relationship with road collision hotspots in that they both present classic examples of geographic phenomena with uncertain boundaries (Longley *et al* 2005: Chapter 6).

Arguably the most important criterion for determining the most appropriate density surface is the bandwidth (Bailey and Gatrell 1995; Brunsden 1991; Brunsden 1995; Fotheringham *et al* 2000) 'Most appropriate' bandwidth applies to finding the suitable measure depending on the dataset and the scale of the dataset. A number of methods to work out the optimum size have been suggested in the literature. Two methods will be discussed here, the first being ESRI's product defaults, whereby the minimum dimension (x or y) of the extent of the input theme is divided by 30 – min (x, y) /30. Bailey and Gatrell (1995) suggest a bandwidth defined by 0.68 times the number of points raised to the -0.2 power scaled to the areal extent of the study area, or $0.68(n)^{-0.2}$. This can be adjusted depending on the size of the study area by multiplying by the square root of the study area size. The problem with both of these methods for estimating bandwidth is that neither takes into account the spatial distribution of the points. Bailey and Gatrell's (1995) estimate is based on point density, but this is limited. Large sample sizes will result in small bandwidths, while small sample sizes will result in large bandwidths. No consideration is given to the relative spacing of the points. The arbitrary nature of the coefficient and power is also problematic. A very large number of combinations would yield similar results. A more practical approach to selecting a bandwidth would take into consideration the relative distribution of points across the study area. One way to achieve this is to base the bandwidth on average distances among points.

It is clear therefore that there are a few guidelines for choosing a particular bandwidth other than by visual inspection (Venables and Ripley 1997). Some have argued that the bandwidth should be no larger than the finest resolution and others have argued for a variation on random nearest neighbour distances (Chainey and Ratcliffe 2005). Others have argued for particular sizes, two methods for which have been outlined in the previous paragraph. Generally, a narrower bandwidth interval leads to a finer mesh density estimate, whereas a larger bandwidth interval will lead to a less clear pattern of variability and therefore less variability between areas. While smaller bandwidths show greater differentiation between areas one has to keep in mind the statistical precision of the estimates. For example, if a sample size is not very large then a smaller bandwidth with lead to greater imprecision in the estimates. On the other hand if the sample size is large then a finer density estimate can be produced. This has entailed detailed experimentation with varying bandwidths in order to find the most suitable. As the with the town centres project (Lloyd *et al* 2005), the final bandwidth is based on a strategic measurement associated with the road network.

Finding the optimum parameters (bandwidth and cell size) for kernel density estimation when analysing road collision density specifically is not an easy process, as there are no strict statistical guidelines which can be followed. The limited range of studies which have documented parameters for road collision density measurements means that the process of deciding the bandwidth and grid cell size is somewhat subjective. In retrospect even if previous research had suggested viable parameters, it is evident that in each study, the area being measured will vary. For example Flahaut's (2004) density measures are of one road in Belgium (N29). Flahaut uses kernel density estimation to estimate the density along this busy and dangerous road in Belgium (Flahaut *et al* 2004). This particular method uses the method implemented by Gasser (1991) which has been subsequently successfully applied to road collision analysis in New Zealand (Sabel *et al* 2005). Both of these studies have very different study area sizes and use a bounded road network for the basis of the analysis.

5.5.2 The selection method and outcomes

The table below (Table 5.1) shows the varying parameters which have been tested and used and the subsequent images of this in ArcGIS. The figures in the table need to be interpreted with caution, as a large majority of the negative (zero readings) where collisions do not occur, fall outside of the study area which is the boundary of the London boroughs where no data have been recorded for this study (the density function creates a square over the study area whereby the majority of negative readings have to be disregarded). Therefore it is important to create quite a fine mesh (in terms of cell size) over the data in order to have a better understanding of the areas where no collisions occur, because if the cell size is too large the cells where there are no collisions would be overlooked as the majority of the cells would contain at least one road collision because of the large number of collisions being analysed. The maps below (Figures 5.5-5.12) show the basic patterns produced by using varying bandwidths and cell size.

Bands	Search radius	Cell Size	m/km	Total no. cells	No. + cells	No. – cells	% +	% -
Band_1	750	250	M	48510	27884	20666	57	43
Band_2	1000	500	M	12180	7335	4845	60	40
Band_3	500	100	M	303450	160459	142991	53	47
Band_4	100	20	M	7586777	1946099	5640978	25	75
Band_5	500	200	M	76007	40123	35884	53	47
Band_6	500	250	M	48510	25671	22839	53	47
Band_7	600	300	M	33775	18551	15224	53	47
Band_8	500	225	M	60138	31717	28421	53	47
Band_9	200	200	M	76007	30500	45507	40	60
Band_10	400	100	M	241120	103414	137706	43	67
Band 11	200	100	M	303450	121915	181535	40	50

Table 5.1: *This table demonstrates the variations in search radius and bandwidth using ESRI's ArcGIS density measure. The negative value in the last column represents the cells in which there were no road collisions.*

Figure 5.5: *Study area, major crossroad in Edmonton, London UK (for a Kernel Density Estimation map for London see Appendix)*

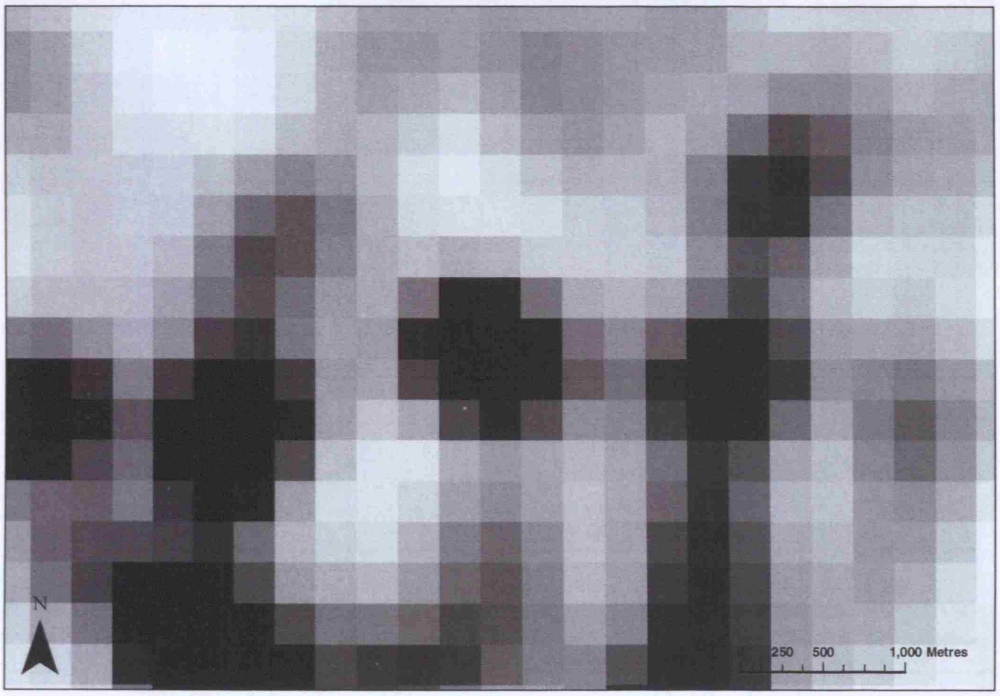


Figure 5.6: *Band 1*

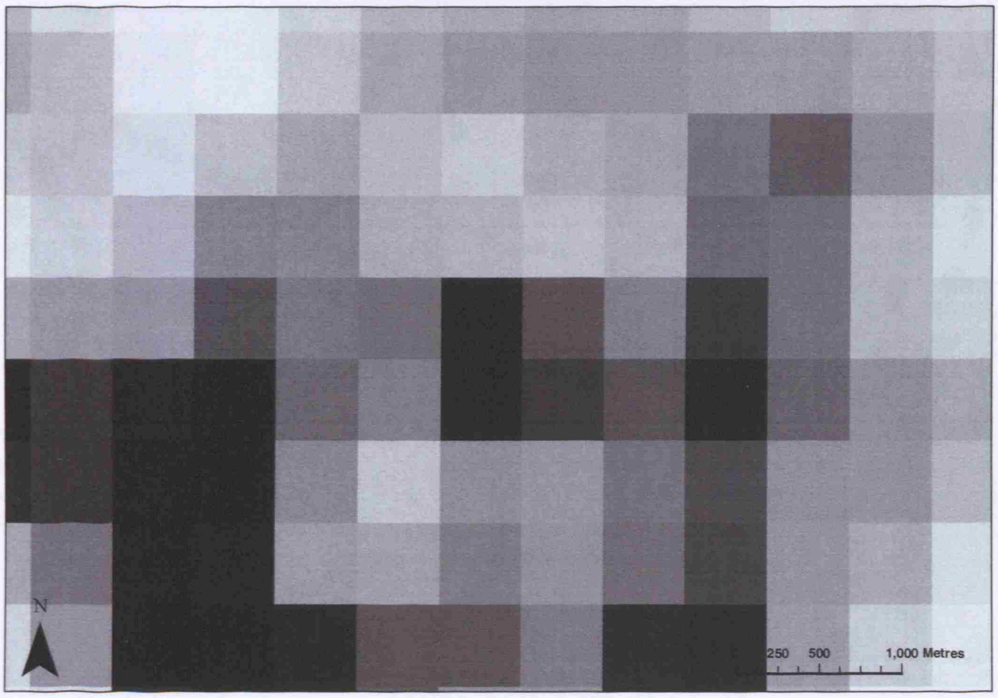


Figure 5.7: *Band 2*

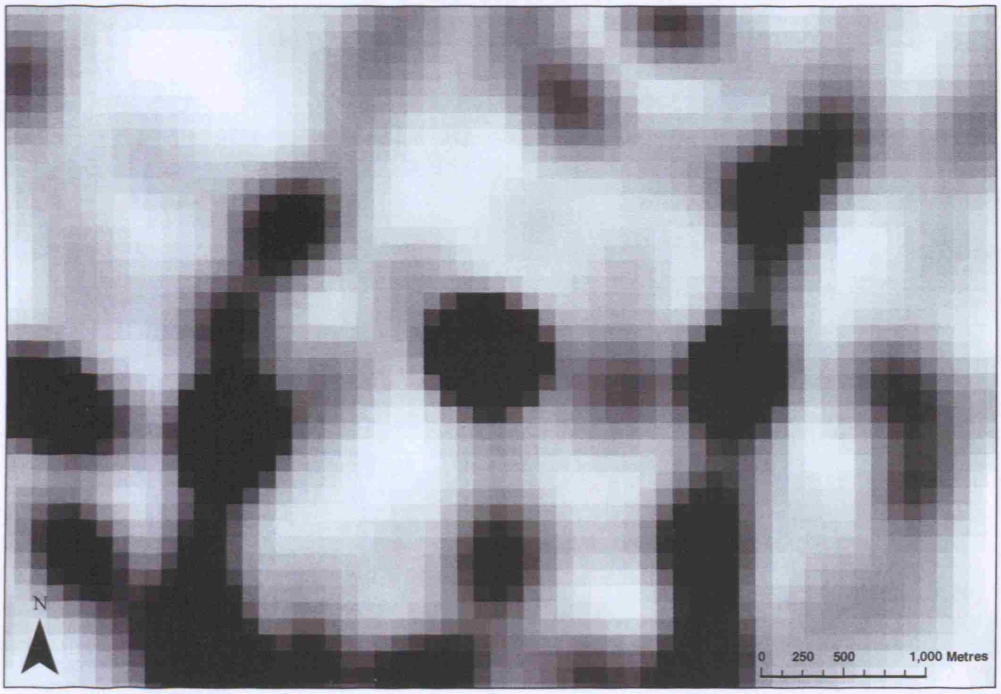


Figure 5.8: *Band 3*

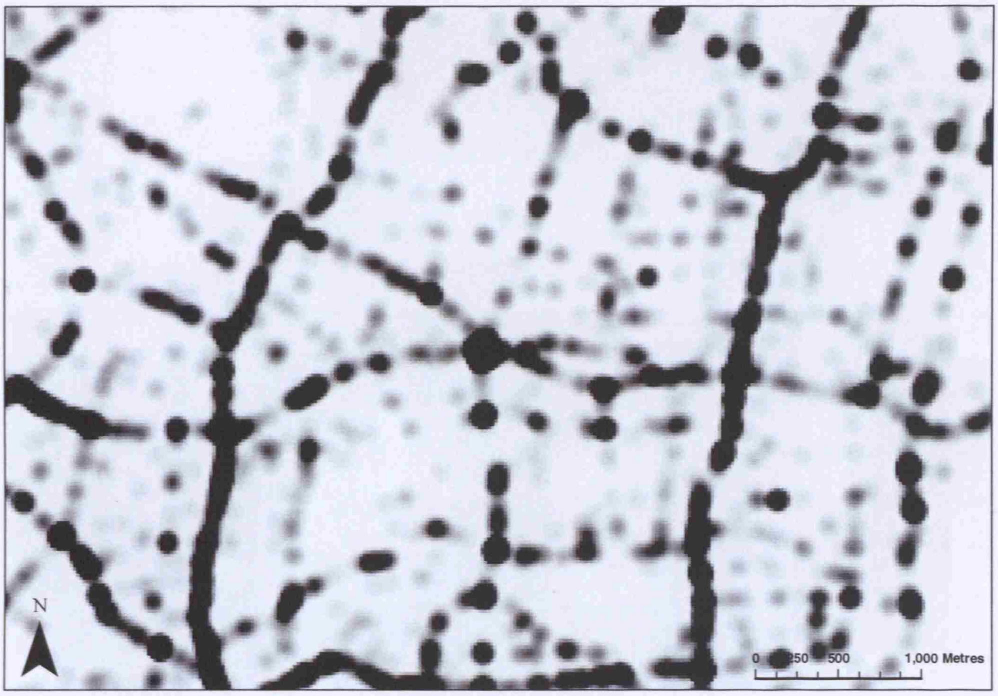


Figure 5.9: *Band 4*

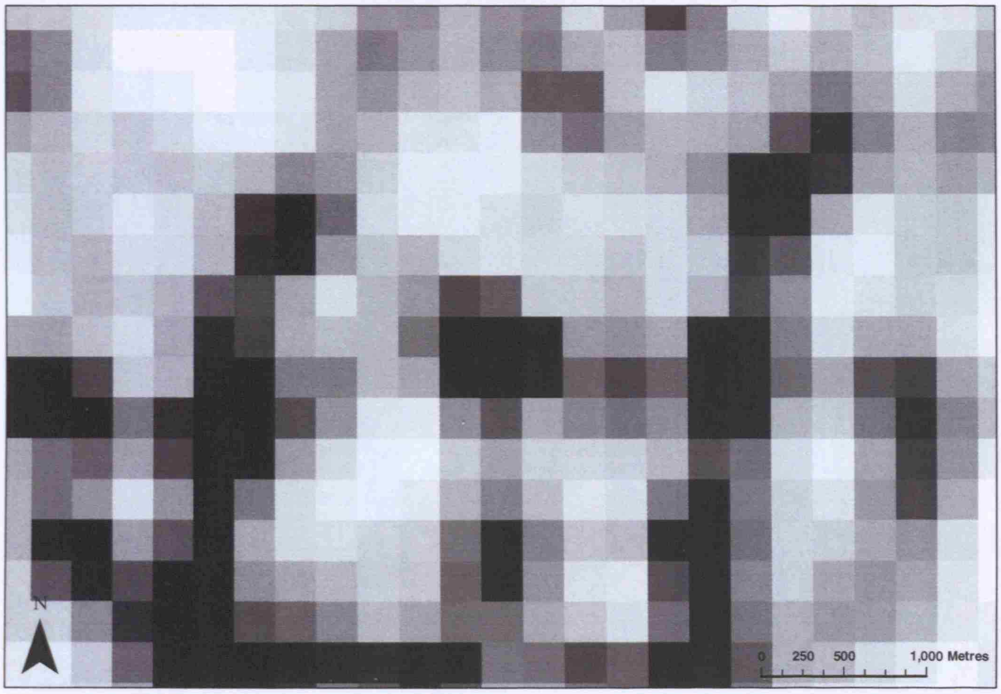


Figure 5.10: *Band 6*

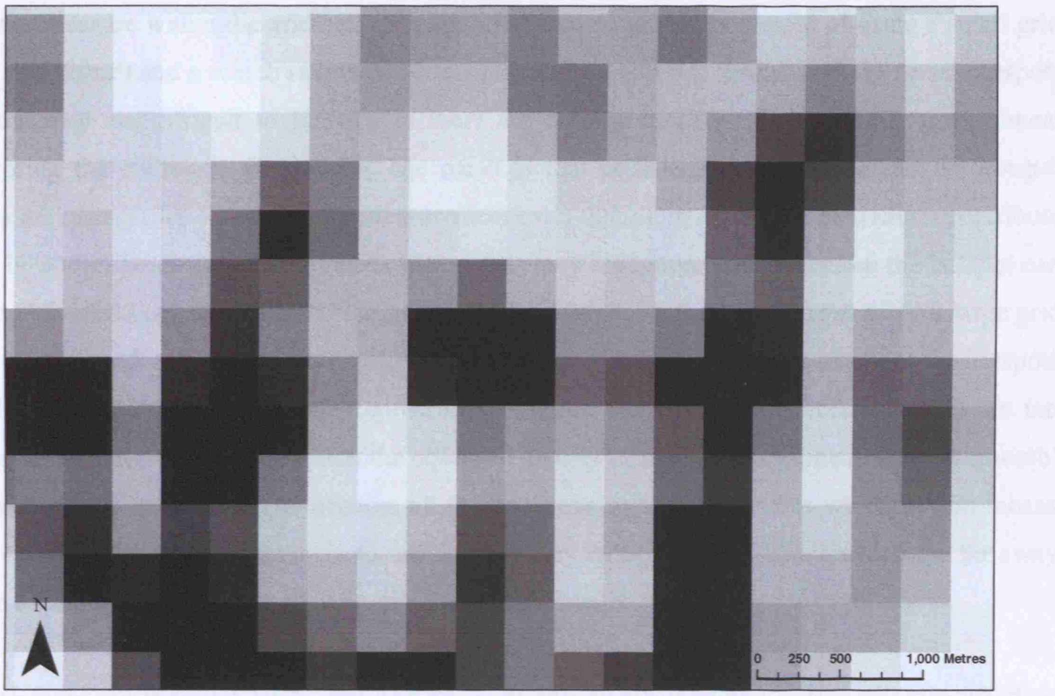


Figure 5.11: *Band 7*

Figure 5.12: *Band 11*

The figures show how changing the bandwidth and grid cell size affects the output of the actual density measure within the grid cells. Figure 5.9 (Band 4) shows the output of using a small grid cell size (20m^2) and a search radius of 100m. The surface appears 'spiky' and the actual hotspots on the map are difficult to identify as there are a large number of very small areas almost mirroring the collisions themselves. The result is that collisions occurring around the hotspot areas are disregarded in the density measurement even though in real space they might contribute to a hotspot. The choice of parameters means that they are ignored and therefore the hotspot can become misleading. In comparison, Figure 5.7 (Band 2) shows the output from using a large grid cell (500m^2) and a large bandwidth (1000m). This map shows an inaccurate image of the hotspots (cells which are darker in colour, representing a higher density) using parameters that are too large in magnitude. This means that the observed density of the hotspots appears as 'too smooth' whereby the method takes into account all the collisions in a 1000m radius which in turn means that the hotspots are too inaccurate to use because they incorporate collisions which are far away from each other and unlikely to have any spatial dependence.

These maps use the same study area and have varying search radii and output cell size which is evident in the images. They illustrate how, in order to determine the optimum search radius, it is necessary to find a surface which is not too fine in granularity, such as that shown in Figure 5.9. At the same time it is important that the surface is not too smooth, whereby only regional patterns would be observed in the results. The surfaces (Figures 5.6-5.12) highlight the results of such parameters. The study area of Edmonton in North London is an area (particularly the crossroads in the maps) which over the past five years has experienced a high propensity of road traffic collisions. The static study area was chosen to portray the differences between the surfaces and assist the aim of finding a suitable bandwidth and cell size. Ultimately these surfaces indicate the different levels of smoothing.

The search radius was set between 100m-1000m and the output cell size between 20m-500m. Band 2 shows the largest cell size (500 square metres) and a bandwidth of 1000m. From the surface it is clear that the cell size is too big (compared to the other surfaces of Band 1 for example) and masks potential hotspots. From the surface results it is apparent that the smaller the search radius (in relation to the cell size) the smaller the spread of risk (in terms of the influence of the collisions over space). The risk containment means the influence of the collision is restricted over space and therefore the search is limited to (in the example of band 9) the extent of the cell (200m). It has been determined that the possible spatial dependence of collisions in an area means that the area around the already identified hotspot may also have an element of risk of a collision occurring. This spread of risk is an important element when trying to prevent collisions in the future, and by addressing and measuring these buffer areas that surround the hotspot itself it means that potential risk can be quantified and managed accordingly.

An objective is therefore to determine the optimum search radius in order to depict the spread of collision risk. The search radius for this purpose has been subjectively tested with a range of distances (see Table 5.1). From the surface images, the changing grid cell size is clearer in the maps than the changing search radius. For example Figure 5.9 (Band 4) shows the search radius and grid cell both constant at 200m. This result is a lot different to the other surfaces (except Figure 5.9: Band 4) whereby the search radius has been kept confined to the size of the cell and the spread of risk around the cell has not been measured. However, because the study area of Greater London (which measures 1620 square kilometres) the search radius has to be small enough not to overlook possible hotspots. However a surface such as Figure 5.9 (Band 4) would be inappropriate as the resulting surface shows a coarse surface with the total number of cells reaching nearly eight million and the proportion of cells which have a density measurement is only 25%, which is well below half and unacceptable for a collision taxonomy.

The final choice was based on taking a search radius which is two times the size of the grid cell, therefore for this study the bandwidth is 200 metres and the grid cell size is 100 metres. Varying the bandwidth alters considerably the number of cells which have a non zero density value. The final bandwidth (Band 11 – Figure 5.12) created 121915 positive density cells and 38676 zero density cells for the Greater London area. This creates a surface which identifies the cells with no density more clearly than cells with a positive density measure. The choice is also based on the road network on which the collisions take place. The average length of road segment (as outlined by OS Mastermap's ITN layer) is 91 metres long, which suggests using a rounded up grid cell size of 100 square metres has relevance to the road network. The average link length for Inner London is 79 metres, compared to 96 metres, as the average link length for Outer London. The average road length in Inner London therefore is 17 metres less than Outer London. However it was decided that a 'rounding up' to 100m would be used as the final choice.

5.6 Limitations of linear surfaces

The kernel density estimation process outlined in this section (using ESRI's ArcGIS) is an area wide smoothing technique. Road collisions are fundamentally constrained to the transportation network and the density process does not take this into account when creating the density surface. The literature that analyses road collisions on a network tends to focus upon the 'optimum segment length' of road to use for hotspot analysis (Thomas 1996). This thesis uses the standard density process, but it is important to be aware that the collisions are constrained spatially when interpreting the method and results for this study.

By using an area wide grid cell spatial unit to capture the hotspots it was possible to link the hotspots causes to land use features and junctions. Using the grid cell also enabled the user to identify both linear (network) and area/zone hotspots. A study in 2003 to identify pedestrian crash zones in Nevada (Pulugurtha *et al* 2003) uses kernel density smoothing to identify two types of spatial collision hotspot. The first type of roadway linear segments (minimum 22 crashes per 2 mile segment) and the second is circular zones with a 300 feet radius. An important point is that although road collisions are constrained to a linear road network, the spatial pattern of high density collisions is both linear and area wide. The grid cell in this context gives patterns of both linear and area hotspots. In the following chapter the spatial configuration of hotspots are clearly linear (particularly large roads in Outer London) and area wide (the centre of Camden near the tube station is an example of this).

The size of the grid cell plays a fundamental role both in determining the initial surface and the end result of the analysis. In short, the smaller the grid cell the larger the number of grid cells that have to be categorised and measured. The number of cells has to be chosen by taking into account the size of the study area. For example Steenberghen *et al* (2004) used a grid cell size of 20m x 20m, which is ten times smaller than the one chosen for this study. The reason for this choice of smaller bandwidth in this particular study arose because the study area was a small town in Belgium with 75000 residents (compared to London's 8.2 million recorded in the 2001 Census). From this evidence, it suggests that a larger grid cell is required in order to create a London-wide study. With respect to scale if analysis was to be done on a disaggregated scale such as at creating a taxonomy at borough level then both a smaller bandwidth and smaller grid cell would be used to take into the smaller study area size.

5.6.1 Using a single search radius

In theory using a different search radius for Inner and Outer London would seem logical given the difference in road density and traffic flow. Inner London has considerably fewer road links (57580) and on average these links are 17m metres less than Outer London which has 129808 road links and an average link of 96m. Although Inner London is a high density hub of traffic flow, similarly in Outer London, areas in Croydon, Lewisham and Clapham experience increased traffic flow and consequently road collisions. Between 1998 and 2003, boroughs with high collision counts included Ealing (7998), Croydon (7205), and Barnet (7568) compared to boroughs such as Islington (6321) and Kensington and Chelsea (5092). Predominantly these boroughs have higher speeds; larger roads and increased traffic flow, in comparison Inner London boroughs have a high proportion of cyclist and pedestrian collisions. Therefore, using a single search radius would make a more uniform pattern of hotspots accounting for the spread of collisions across London.

To estimate the density of each 100m^2 raster cell across London the following process was used. The requirements were to have spatially referenced kernel density grid cells and a corresponding z-score representing the density.

- A raster surface was created for all collisions in London using the density function in ArcGIS Spatial Analyst, using the appropriate bandwidths and cell size.

- This raster surface was then exported as an X and Y coordinate of each cell centroid and a Z value representing the density. The surface was exported using the Raster Calculator function in Spatial Analyst using the equation below:

  Outfile.txt = SAMPLE [(Ingrid), (Ingrid)]

  Ingrid = file name of surface

  (Source: ESRI Technical Notes 2005)

- The output file is a text file which is then exported to Access (because of the high number of records) and is divided into positive and zero Z values. The positive density records are then transferred into Excel where the score is multiplied by 1000. This gives a 'true' reading for the density score and equates for the size of the cell, and helps identify the high scoring cells. It has been identified in the literature as a 'scaling up' factor which enables the user to convert the very small numbers generated by KDE to easier numbers to manipulate and use.

- The X and Y centroid file is then mapped in ArcGIS. Using a downloadable spatial tools extension (www.spatialecology.com 2005); it was possible to create a polygon grid over the raster surface. Therefore each extracted density point has a grid cell surrounding it.

- It was then possible to select the grid cells using the density points which were over a certain level. The threshold in this case was 1, because it represented a significant number of cells which had a high density. The scaling factor was 1000 because the Z values had four decimal places and this would make whole numbers from the results.

- The resulting surface is a selection of grid cells (2290) which have a significantly high density. These grid cells are scattered evenly across London and a large proportion are grouped together indicating differing sizes in high density hotspots.

5.7 Delineating hotspots for analysis

5.7.1 Determining the sites for further statistical analysis

All cells which have a certain level of density will be investigated further and used to create an accident typology. In other words, the higher the density measure within a cell the more collisions are likely to occur in that area (based only on historical data). The ranking of sites has been much neglected in the literature and the aim of this section is to determine a suitable ranking system with which to order the hotspots according to accident density. To some degree the methods will inevitably be subjective, as there is limited and varying framework, guidelines or literature which can be used as a proxy indicator as the most suitable thresholds. Some local authorities have their own individual methods and frameworks for hotspot thresholds and ranking the most important and treatable sites. This makes comparison across a London wide area almost impossible, which is why for the purpose of this study a unique London wide threshold framework has been devised in order to make meaningful comparisons across space.

In order to establish a suitable ranking threshold it is important to decide whether a universal ranking order should be used (to accommodate all surface scores across all time periods) or to create a threshold for each individual time band and severity surface. Possible methods for ranking have been mentioned in the literature review for this section. From this review it is evident that there is no correct ranking procedure and that ranking alone might not produce the most robust results. If each surface were to have its own ranking then it would be challenging to compare the results between the surfaces; however this in turn compromises the ranking system whereby for example in time band one the dataset is smaller than the others and hotspots may be overlooked as the threshold may be too high. These are important considerations when determining a threshold framework within which to base the statistical analysis.

The other major consideration to bear in mind is the choice of threshold and whether or not it is more suitable to have one threshold above which all cells are subjected to further analysis or whether a scaled ranking method should be used, whereby the hotspots are ranked in order of density and a range of cells are statistically investigated.

The hotspots which have been allocated for further investigation are those which fall above a certain density level. The next stage of this process was to select the cells which fall above certain threshold to analyse. The cells which fell above this threshold were selected. Often a number of these cells were adjacent to one other which meant that the total number of cells which were contiguous were selected and defined as a hotspot. The collisions within these cells were then selected for further analysis.

5.8 A surface based approach: an appraisal

Road collision point pattern analysis has derived a large proportion of its techniques from crime pattern analysis. Kernel density estimation has become one of the more robust and appropriate techniques used in mapping patterns of crime (Chainey *et al* 2002). Kernel density estimation has become more widely used in road collision analysis for its detection of high density collision areas, or hotspots. Its aim is to give an underlying visual impression of the patterns associated with the events. This chapter however has endeavoured to go beyond merely a visual interpretation by extracting the density information to form a database of spatially referenced hotspots and a density score. By creating this database of spatially dependant collisions it will begin the next stage of understanding the spatial patterns and causal factors associated with these hotspots.

These hotspots were chosen because they had a density which was above a particular threshold. In addition to these 428 hotspots, there is a high proportion of hotspots which falls below this threshold yet still pose a road safety risk. Although this is not going to be investigated it is important to remember that these hotspots in this study represent the highest density hotspots within London. Similarly, there are a number of grid cells in London which have a density value of zero indicated little or no collisions in these areas. It is possible that some pattern exists which influence the under representation of collisions in these areas, and this would be a potential research strand to explore in future studies.

There have been few studies in the road safety research domain which have used kernel density smoothing and extracting the data from the surfaces. A recent study in Christchurch, New Zealand (Sabel *et al* 2006) has shown the feasibility and success of extracting a road collision density surface and traffic flow data in order to depict hotspots and using Monte Carlo simulation, to try and predict the location of future hotspots. The research outlined in this chapter has used an area based approach using a raster grid cell as the basic spatial unit of the hotspot. In the output of the hotspot surface, it was clear that certain hotspots were route, area or junction hotspots from the shape of the grid cells. Sabel (*et al* 2006) uses a combination of both kernel grid cells and a linear road network feature to depict road collision hotspots against traffic flow. Using a grid cell approach has the added advantage one is able to define area as well as linear hotspots, an important tool when nearly ¾ of London's collisions occur at junctions or within 20 metres of them. This chapter has presented an innovative methodology for detecting road collision hotspots. It presents both a theoretical framework for kernel density estimation choices for road collisions and a step by step technical guide. The next chapter explores the characteristics of these hotspots, to create a clustering model.

CHAPTER 6

PROFILING THE ROAD COLLISION HOTSPOTS

Section A

6.1 The clustering process

6.1.1 Introduction

The classification of road collisions currently used by Transport for London road safety initiatives typically orders collisions by severity, or by involvement of vulnerable user groups such as a cyclists or pedestrians. These criteria are used as the basis for road safety research and counter measures. For example the 'Teen Road Safety Campaign' in Central London has brought increasing awareness to teenagers to the risks of the road environment through a series of adverts and posters within London (Transport for London, Street Management 2005). In conjunction with this are the criteria for selecting road safety speed camera sites which are located where four or more fatal or serious injury collisions have taken place. It is a widely used GIS technique for example to extract the collisions which are fatal and map them and try and detect any spatial pattern to target. If a significant number of fatal collisions are found close to one another, it can be inferred there may be a factor which links these collisions together. As we have seen in the previous chapter the method of determining collision hotspots can be very complex and can go further than disaggregating the data. There is a clear divide operationally between the spatial and non spatial analysis of road collisions. The non spatial refers to the analysis of the collision data set, the percentage of cyclists in road collisions, the number of pedestrians. The aim is always to find areas where there are many collisions occurring within a small distance of one another. This is a basic cost benefit approach and is used for collision counter measures perhaps of an engineering nature. However if our objective is to understand the social reasons rather than the engineering factors we need to use different methods whereby the nature of hotspots and method of defining them is far more complicated than dots on a map. This does not represent the nature and spatial pattern of the collisions accurately. Therefore by using the density measure from the previous chapter, in this chapter we show how hotspots can be described using an approach more appropriate for understanding of the social reasons behind different types of collisions in different spatial locations.

There are multiple reasons for many collisions – a particular set of collisions may occur at a location on account of the fact that many pedestrians in that locality emerge from clubs late at night in a drunken stupor. But they also may occur because there is not a barrier separating pedestrians from fast moving traffic. The problem of the collision can be addressed both in terms of the social reasons and the engineering reasons. It might reasonably be argued that the most effective engineering solutions at a particular location are difficult to identify without some consideration of the immediate social context – for example having warning notices to pedestrians, an engineering solution, may not be as effective in an environment where the accidents happen to late night revellers than in an environment in which the principal hazard is tourist shoppers.

Therefore the reasons for there being so many collisions in one hotspot may vary from hotspot to hotspot. For example in one hotspot it may be due to the increased interaction between cyclists and traffic at a certain time of day. Or, it could be due to the limited visibility for vehicles to see pedestrians crossing the road. Both these hypothetical hotspots will ultimately result in different measures to reduce the number of collisions in the hotspot. Hotspots therefore may have the same pattern of causation and why there are so many collisions in one hotspot. This would result in similar measures being taken for a number of hotspots in order to reduce the number of collisions. Therefore, by conducting this clustering it will be possible to determine the similarities and differences between hotspots in order to understand why they occur and the potential treatment methods.

The methods used in the previous chapter created a surface whereby the grid cells represent the hotspots based on the density measure. The number of grid cells in a hotspot varies, showing the hotspots are not uniform in size or shape. To be part of a hotspot a grid cell has to have a collision density level which is over a specified threshold, indicating these are the areas in London where the collision density prevalence is at its most intense. This grid surface provides the basis for collating the collisions which occur within these grid based hotspots.

The product of this methodology is a database of hotspots of varying size and density and associated collisions which fall within the boundaries of the hotspot. The nature of the database means that there is no analysis of individual collisions, but an analysis of the set of the collisions which share a common nearby spatial location, implying a common and linking casual factor. Although hotspots may be treated as though each is unique, they may share similar characteristics such as the proportion of pedestrians or cyclists or an increased number of collisions in certain weather types or they may occur at a certain time of day or particular day of week. By ascertaining the nature of this similarity, comparisons between hotspots can be made on a 'like by like' basis.

This chapter aims to address the nature of the similarity between the hotspots using a clustering method developed by Professor Richard Webber. The algorithm clusters the hotspots and associated attributes and aggregates the data into a number of clusters, based on the attribute data and level of similarity. Each hotspot, and hence each collision within is then be assigned to a cluster, outlined in the diagram below:

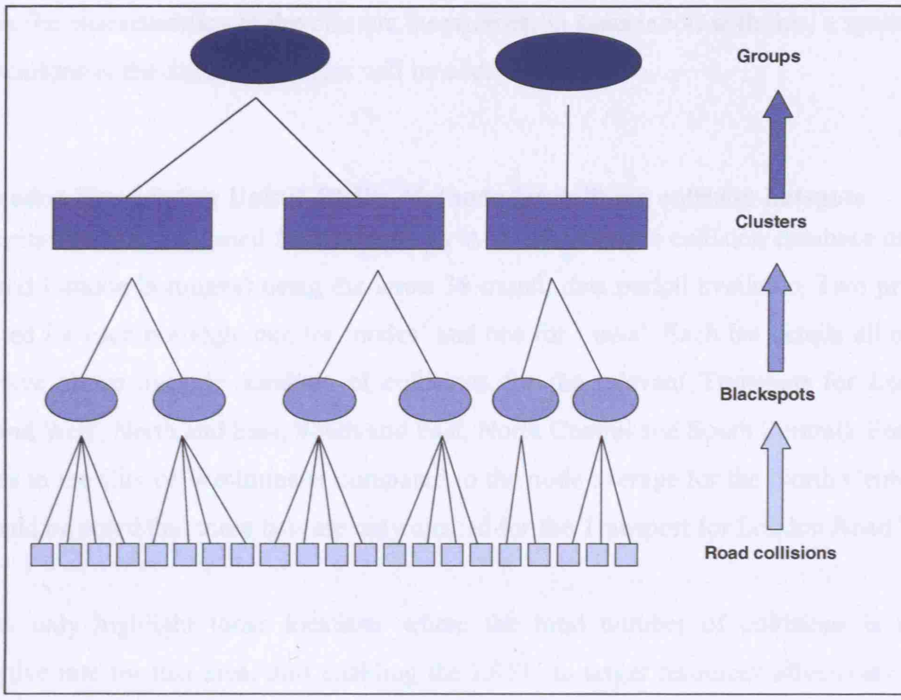


Figure 6.1: *How the hotspot classification method works (Anderson 2006)*

Figure 6.1 shows how at a primary level, collisions occur on the road network. The collisions which share a spatial commonality within the grid cell are selected. These hotspots are then classified using a clustering process and are organised into classes (or clusters) based on similar attributes. These clusters are then organised into groups, based on the similarity of the clusters. This hierarchical process allows a classification of spatial hotspots based on similarity of either the characteristics of the collisions within the hotspots or of the environmental and land use information associated with the hotspot area itself or indeed of both. This guarantees a more rounded approach to the nature of the factors associated with causing a high number of collisions in such a small area.

The object of the clustering methodology is to identify the similarities and the characteristics of the hotspots and to outline in detail the nature of the similarities which occur within the data. The first section of this chapter explores an example of how hotspots are defined by the London road safety unit. It continues by describing the clustering methodology adopted for this particular study, the nature of the data, including normalisation, weighting and choosing the number of clusters. The final part of the first section looks at the limitations associated with the methodology. The second section reviews and analyses the outcome of the clustering results and identifies the characteristics of the clusters themselves. In association with this, a spatial analysis of the locations of the different clusters will be addressed.

6.1.2 London Road Safety Unit (LRSU): Methods for defining collision hotspots

The priority lists are generated from reports by ACCSTATS (the collision database used by the LRSU and London boroughs) using the latest 36 month data period available. Two priority lists are created for each borough, one for 'nodes' and one for 'links'. Each list details all of the sites which have above average numbers of collisions for the relevant Transport for London area (North and West, North and East, South and East, North Central and South Central). For example, the nodes in the City of Westminster compared to the node average for the North Central area. It also should be noted that these lists are only created for the Transport for London Road Network.

The lists only highlight those locations where the total number of collisions is above the comparative rate for that area, thus enabling the LRSU to target resources effectively. The sites are then prioritised. For the 'nodes' priority list, the sites are ranked according the total number of collisions. For the 'links' priority lists, the sites are ranked by the calculated number of collisions per kilometre, the figures for which are also displayed on the lists.

For both the 'links' and 'nodes' priority lists, in addition to the total number of collisions over the 36 month period, the annual totals are also displayed to highlight those sites which have revealed a year on year increase in the number of collisions. For each location the total number of collisions is disaggregated to identity the main problem areas (considered by the LRSU), which are outlined below:

- Killed or serious injury collisions
- Collisions involving pedestrians
- Powered two wheeled vehicle collisions
- Pedal cycle collisions
- Right turn collisions
- Collisions during the hours of darkness
- Collisions on a non dry road surface

Cells in the spreadsheet are highlighted blue to indicate where the proportion of the number of collisions is greater at that site than the average at similar sites in Inner or Outer London, therefore giving an indication of the collision problems to be found at that site.

6.1.3 Data input – sources of data for hotspot classification

When determining the database used to build the classification it is important to assess the type of data which would be collected and would have the potential of having impact on collision density. Therefore it was important to consider not just the attributes of the collisions themselves but environmental and land use data which were found in the vicinity of the hotspots. Table 6.1 shows the chosen attributes and associated source.

Attribute	Source
Road length	Ordnance Survey Mastermap™
Cycle lane length	London Cycle Network
Pedestrian Crossings	Transport for London
London Underground stations	Transport for London
Traffic lights	Transport for London
Bus stops	Transport for London
Schools (primary and secondary)	Department for Education
Speed cameras	London Safety Camera Partnership

Table 6.1: *Environmental and land use data and associated sources*

Some of the data varied in format. For example the network data (road and cycle) were in polyline format and their total length had to be established for each hotspot. The pedestrian crossing data were in the format of a geographically referenced 2-D shape, representing the exact location and outline on a map. The remaining data were in point format, which meant the count number of each variable could be calculated. All the data has been collated into one single geography, the basic spatial unit of the hotspot which is the 100m^2 grid cell.

To select the hotspots, various rules were established to make the process simpler. Hotspots were established by the linking of cells. A hotspot could be made up of one single cell of many cells. The rule was that the cells had to join sides, and no diagonally linked cells would constitute being part of the hotspot. This was a difficult decision to make, as there is a valid argument for cells linking diagonally might be spatially linked. However, if the cells are similar then this should be distinguished in the clustering outcomes.

The collision data added into the hotspot database were derived from the STAT19 database in which all the data were in count format, relating to the collisions occurring within the hotspot. For example, the severity of each collision (fatal, serious or slight) within the hotspot is added to the database. Each of these counts including the added non STATS19 counts needed to be related to a corresponding base count. The main objective in creating a logical classification is to create an accurate representation of the data; therefore the data must have a suitable basis for comparison.

The data in this instance have been normalised by one of two variables, the number of collisions in each hotspot or the number of grid cells that make up each hotspot. The reason for this, was to accommodate the two different types of data being used (STATS19 and environmental/road network). The Stats19 collision type counts were divided by the number of collisions within the hotspot, and the environmental/road network data were divided by the number of grid cells as these data were related to the area within the grid cell.

	Original data			**Normalised data**		
Hotspot ID	Bus stops	Severity = slight	Pedestrians	Bus stops (by cell)	Severity = slight (by collision)	Pedestrians (by collision)
Th7	63	97	27	9.00	0.85	0.24
Gre13	6	29	4	6.00	0.76	0.08
Cro7	8	30	16	4.00	0.94	0.16
Cro1	3	26	7	3.00	0.76	0.06
Cro3	7	43	15	2.33	0.84	0.08

Table 6.2: *Original and normalised data for hotspot database*

Table 6.2 shows the data in its original count format and final three columns show the data when they have been normalised. For example the number of bus stops in hotspot 'Th7' was 63 and the normalised value was 63 divided by the number of grid cells within the hotspot (7 grid cells). The columns for severity and pedestrians were normalised against the number of collisions within the hotspot.

6.1.4 Weighting the variables

For the purpose of the clustering methodology the variables have to be given weights which will make them more or less influential in the cluster outcome. It is important to give careful consideration to the appropriateness of the weights given to the different variables. The variables have to be assessed when deciding the weights for each variable due to how strong an influence it was desired that the variable should have on the outcome (this is equivalent to inputting the variable many times). For example all the time variables were treated equally, similarly the 'day of week' variable were also treated equally. This meant that the clustering process would give equal weight to all those variables. The exception occurred for example to severity, whereby fatal collisions occur less frequently than slight ones and therefore if the same weight had been given too all three severity types the results would have been skewed because there is a considerable more likelihood of being involved in a slight collision rather than a fatal collision. Therefore the weighting for slight collisions was higher compared to fatal collisions. The clustering programme uses weights 1-10, 10 resulting in the variable having ten times more influence than a variable with a weight of one. Therefore the table below outlines the selected weights for each of the variables being entered into the clustering programme. One reason for varying the weights is that there may be more variables relating to one aspect of a collision's character than another. For example time attributes may be more plentiful than weather attributes, if all the variables were given the same weight the solution would be skewed towards those dimensions for which there are more variables.

Variable	Weight	Variable	Weight	Variable	Weight
Hotspot ID	0	Time 4	2	Friday	2
Number	3	Time 5	2	Saturday	2
Collisions	3	Time 6	2	Sunday	2
Number cells	0	Time 7	2	Weekend	5
Speed	5	Time 8	2	Vehicle 1	4
Road Length	4	Time 9	2	Vehicle 2	4
School	4	Dark	5	Vehicle 3	2
Tube	3	Casualty 1	4	Vehicle 4	2
Cycle length	4	Casualty 2	4	Vehicle 5	1
Traffic	4	Casualty 3	3	Average	4
Bus stops	5	Casualty 4	2	Weather 1	2
Pedestrian	4	Casualty 5	2	Weather 2	2
Slight	6	Casualty 6	1	Weather 3	2
Serious	3	Average	5	Weather 4	2
Fatal	2	Monday	2	Weather 5	2
Time 1	2	Tuesday	2	Weather 8	2
Time2	2	Wednesday	2	Weather 9	2
Time 3	2	Thursday	2	Pedestrians	5

		Cyclists	5

Table 6.3: *Variables and associated weights*

Table 6.3 shows all the variables used in the clustering procedure. The different weights highlight the differences in the relative importance of certain variables within the procedure. For example, the proportion of slight collisions is given more weighting than fatal. This is because if there was a fatal collision within the hotspot, it would be over represented because it is such a rare event, therefore to balance out the weights, and slight collisions are given more influence. This is also because there are more slight collisions and therefore statistically this is a more reliable variable. The variable day of week is given equal influence but the weekend variable is given more weight as it is a combined variable from two others, therefore combining the importance of two existing variables. The environmental and road network variables were given no weighting. They are used not to influence the classification but rather to interpret the categories and to establish the relationship between the categorisation and environmental factors. In two test runs of the clustering programme the first weighting the environmental or non collision variables and the second run not weighting them, there was a significant difference in the outcome. Weighting the environmental variables resulted in very high and unlikely index scores in the clusters which distorted the results as they were above average. This pattern was present for nearly all of the environmental variables. The second run which gave zero weight to the environmental variables had an outcome whereby the percentage of variance in the environmental variables was very small, indicating that the variables could not be explained accurately. Therefore it was decided not to weight the environmental variables as they were not predictive of differences in the actual collision data.

6.2 The method of clustering

This procedure represents an attempt to classify each of the 428 hotspots into relatively homogenous types based on their environmental characteristics. The principal benefit of this process is that it generates solutions which account for a higher proportion of the variance of the source variables and clusters which are more equal in population size (Harris 2003). The initial stage of the clustering process starts with specifying the most likely number of clusters in the classification that would be appropriate. It should be mentioned that the diagnostics will indicate whether a suitable number of clusters has been selected. If it has not it is easy to rerun the programme with a higher or lower number of clusters.

Choosing the number of clusters which are created (a 'bottom-up' approach) means that the user has a larger degree of control over the process. After consideration it was decided that fifteen clusters would be determined and these would be aggregated into five groups. Five groups were chosen because this would aim to allocate three clusters to each group making it a divisible number. This gave on average 30 hotspots per cluster and three clusters per group. However, the clustering process is unlikely to produce such uniform between group results and indeed did not in this case.

There are a broad number of clustering algorithms available. This study uses K means, following the procedure outlined in Figure 6.2.

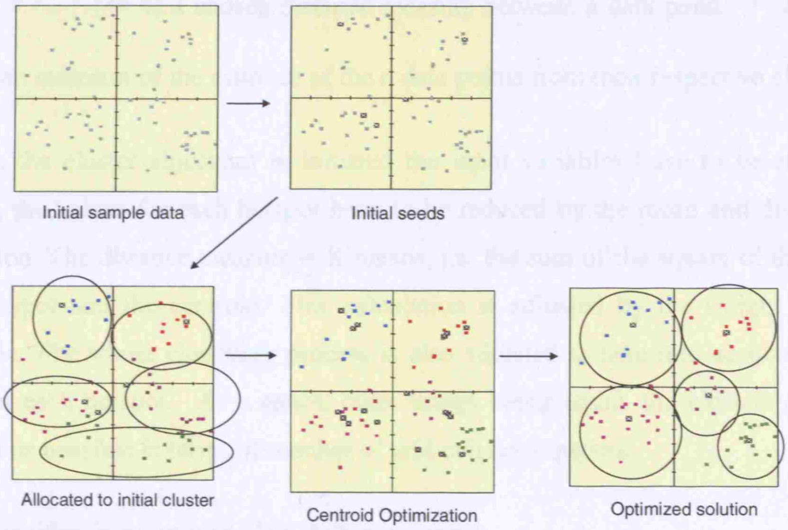


Figure 6.2: *K-Means clustering diagram (Source: Cameron 2004 as cited in Harris et al 2005)*

The algorithm starts by defining K hotspots, one for each of the predefined clusters. These are selected on a random basis but proportional to the number of grid cells in the hotspot. These are used as cluster centroids. The next step is to take each point belonging to a given cluster data set, in this case each hotspot, and measure its degree of similarity to each centroid in terms of values on the input variables, duly weighted. It is then allocated to the nearest centroid. At the conclusion of this process the average co-ordinates of the K means centroids need to be calculated

this being the average of the values of the hotspots allocated to each cluster. This modification of the cluster average is referred to as the iteration. As a result of this iteration some hotspots will now be closer in terms of similarity to a different cluster centroid than those to which they were assigned to at the end of the previous iteration. Thus further iterations are undertaken until the cluster centroids and the allocation of the hotspots to clusters stabilises.
When this process has been completed the algorithm will have achieved its aim which is to minimize an objective function, in this case a squared error function. The objective function:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 ,$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the n data points from their respective cluster centres.

Before the cluster algorithm is initiated the input variables have to be standardised. In other words, the values for each hotspot have to be reduced by the mean and divided by the standard deviation. The distance measure is K means, i.e. the sum of the square of the difference between the hotspot and the centroid. This calculation is adjusted by the weight given of each input variable. The whole clustering process is also adjusted to take into account the number of grid cells in each hotspot. As a result, other things being equal, the clusters that result will be as similar as possible in terms of number of grid cell not hotspots.

The algorithm is composed of the following steps:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

The next stage of the process involved using hierarchical clustering, a method which has been used for this study for a variety of reasons. The fundamental aim in this case of hierarchical clustering is to organise the 15 clusters into five groups. When defining road collision hotspots it is of fundamental importance to determine those hotspots which are of the highest priority for resource allocation. There are many ways to distinguish the priority of hotspots.

Each of the clusters of hotspots begins as an entity in its own right. The first stage is to find the 'most similar' pair of clusters, i.e. the pair that could be combined for the least incremental loss of variance of the original data; these are merged into a parent cluster and the process repeated until all the clusters have been joined together. The similarity between clusters is defined in terms of the distance between clusters weighted by the number of grid cells. The end result is represented by a dendrogram or taxonomy or hierarchy of data points. The next part of the chapter will outline the results from the clustering process and discuss the hotspot classifications in depth.

Section B

6.3 The clustering classifications and analysis

6.3.1 Introduction

The method of clustering road collision data is different to that of say the clustering processes used to create geodemographic marketing solutions. The difference is the unique nature of the data. The uniqueness of road collision data is due to its static nature as it tries to capture a moment in time. This chapter aims to collate the causes and patterns present in a road collision database. Clustering is a method which helps to order the data into a meaningful outcome which can be utilised by road planners, engineers and policy makers alike. The classification process and subsequent mapping will mean that the clusters can be managed on a more 'like-by-like' basis. This classification information is linked to the postcodes of the drivers and the casualties in order to depict the likelihood of what type of collision a person may likely to be involved in based on their postcode. This clustering process will also highlight the importance of cross border partnership with reference to the borough road safety initiatives. Many hotspots occur across boundaries and hotspots have a similarity across London.

The previous section outlined the finer details of the clustering process and the decisions associated with the methodology. This section discusses the outcome of the clustering programme in three stages. The first stage will address the general diagnostics of the results in order to ascertain how successful the method has been. The second stage will address the index scores of each of the clusters and groups and aim to give a contextual description of each cluster based on the scores of each variable. This section also addresses the spatial patterns of the clusters and groups within London and the spatial patterns within individual boroughs. The final section discusses the nature of road safety campaigns within London with respect to the types and groups.

6.3.2 The clustering process outcome

The success of the clustering process is based of what is referred to as the percentage of variance which is explained by the clustering programme. In this instance the variance explained is 34%. Although this appears to be a low figure, previous studies indicate that this figure is an 'acceptable'. A study by Naveh *et al* (2002), who tried to depict the causes of fatal road collisions and injury only road collisions found the variance for fatal collisions, 2% and for injury only collisions, 12%. (One has also to remember than the 34% is the average for all the variables – for some the percentage is higher, for others lower. Therefore from these results a variance of 34% seems significantly higher especially considering the static nature of the collision data.

6.3.3 Variables and explained variance

Different variables account for different amounts of the observed variance in the dataset. This section will discuss the nature of the variance in both the groups and the clusters.

Variable	% Variance explained Groups	% Variance explained Clusters
Number of collisions	82.9	83.0
Number of cells	82.7	82.8
Cyclists	70.0	75.7
Number casualties = 3	71.6	71.8
Casualties per collision	61.8	63.6
Pedestrians	33.6	62.1
Number of vehicles = 1	42.5	57.4
Dark	6.1	55.7
Collisions per cell	52.8	53.2
Weekend	0.4	50.0

Table 6.4: *Highest variance by cluster scores*

Table 6.4 depicts the top 10 variables whose variance was best explained by the cluster process in terms of the amount of variance in the dataset that they explain. In other words there are the characteristics in terms of which hotspots in any cluster are broadly similar. The top two variables, number of collisions and number of cells, are by far the strongest variables and these two variables are fairly stable in the dataset. The variables of most interest are the high variance explained for the pedestrian and cyclist collision and the number of casualties (3) being of significant importance. There is a good variance explanation of weekend collisions and collisions occurring in the dark. By contrast the table below shows the variables with the smallest variance. This does not indicate that these variables are any less robust merely that less variance between hotspots within a cluster is relative large compared between variance between clusters.

Variable	% Variance explained Groups	% Variance explained Clusters
Traffic lights	3.0	7.4
Number of speed cameras	3.4	7.0
Number casualties = 4	4.9	7.0
Tuesday	1.0	6.8
Fine with wind	4.0	6.2
Bus Stops	1.2	5.7
Number of vehicles = 4	3.0	5.2
Number casualties = 6	2.1	4.7
Schools	3.2	4.4
Tube stations	2.9	4.0
Snowing	1.4	3.5

Table 6.5: *Lowest variance by cluster*

Table 6.5 above shows the variables with the lowest variance explained by cluster variable. Of significant is the high number of environmental variables present including tube stations, schools, bus stops, speed cameras and traffic lights.

Variable	% Variance explained Groups	% Variance explained Clusters
Number of collisions	82.9	83.0
Number of cells	82.7	82.8
Number of casualties = 3	71.6	71.8
Cyclists	70.0	75.7
Casualties per collision	61.8	63.6
Collisions per cell	52.8	53.2
Number of vehicles = 1	42.5	57.4
Number of vehicles = 2	42.1	49.2
Number of casualties = 2	40.8	45.8

Table 6.6: *Table of highest variance determined by group*

The variables in this table are ordered by the percentage of variance explained at the group level. These results are similar to those shown in the cluster table (Table 6.6), whereby the cyclist and pedestrian collisions are explained more comprehensively, as with the number of vehicles involved. The table (Table 6.7) below highlights the lowest scoring variables in the groups, showing interestingly a high proportion of time bands including collisions occurring between 5pm and 7pm and those collisions occurring at weekends. Low variance may occur either where we have a very low sample size for that variable or where its distribution is relatively random, i.e. not influenced by the contextual characteristics of the hotspot.

Variable	% Variance explained Groups	% Variance explained Clusters
Time 2100_2359	1.5	30.5
Snowing	1.4	3.5
Wednesday	1.4	9.4
Bus Stops	1.2	5.7
Tuesday	1.0	6.8
Time 1500_1659	0.8	12.1
Unknown weather	0.7	14.6
Weekend	0.4	50.0
Time 1700_1859	0.4	8.9

Table 6.7: *Table of variance by group distribution*

6.3.4 The distribution of groups and clusters

In the clustering process the number of clusters and groups are predetermined by the user. This section outlines the clusters (and counts) and how they are placed into the different groups.

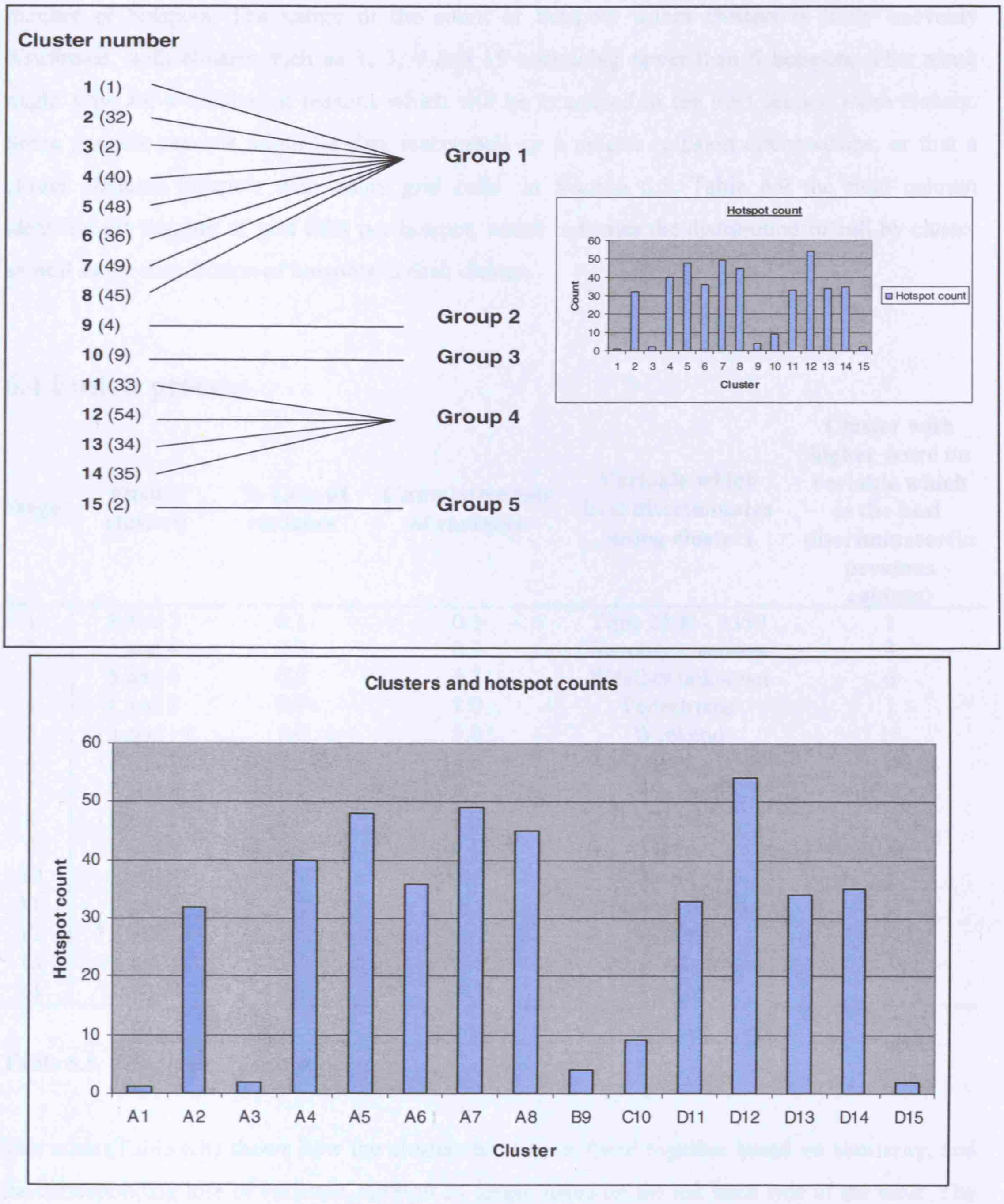


Figure 6.3: *The clusters and associated groups (count in brackets)*

The diagram above (Figure 6.3) shows the number of clusters and in brackets the number of hotspots within the clusters, and then the clusters membership of the groups. The graph depicts the number of hotspots in each cluster. From the diagram it is clear that the group 1 and 4 contain the most clusters and groups 2, 3 and five only contain one cluster, each containing a very small number of hotspots. The nature of the count of hotspots within clusters is fairly unevenly distributed, with clusters such as 1, 3, 9 and 15 containing fewer than 5 hotspots. This result might arise for a number of reasons which will be examined in the next section more closely. Some possible reasons might be data inaccuracy, or a unique collision circumstance, or that a cluster contains hotspots with many grid cells. In Section 6.5, Table 6.9 the final column identifies the number of grid cells per hotspot, which indicates the distribution of cell by cluster as well as the distribution of hotspots in each cluster.

6.4 Fusion process

Stage	Fusing clusters	% Loss of variance	Cumulative loss of variance	Variable which best discriminates fusing clusters	Cluster with higher score on variable which is the best discriminator(in previous column)
1	1 and 2	0.1	0.1	Time 2100 - 2359	1
2	3 and 4	0.2	0.3	Severity = serious	3
3	5 and 6	0.9	1.1	Weather unknown	6
4	1 and 3	0.9	2.0	Pedestrians	1
5	13 and 14	1.0	3.0	Weekend	13
6	11 and 12	1.1	4.1	Time 0000 - 0359	11
7	7 and 8	1.2	5.3	Weekend	8
8	11 and 13	2.0	7.3	Dark	11
9	1 and 5	2.3	9.6	Dark	5
10	1 and 7	2.9	12.5	Weekend	1
11	1 and 9	3.6	16.1	Number of cells	9
12	11 and 15	5.0	21.1	Number of	15
13	1 and 10	5.8	26.9	Collisions per cell	10
14	1 and 11	7.2	34.1	Number of	11

Table 6.8: *The cluster fusion process*

This table (Table 6.8) shows how the clusters have been fused together based on similarity, and the corresponding loss of variance, through 14 stages listed on the left hand side of the table. The second column refers to how the clusters are fused to the most similar, therefore clusters 1 and 2 are similar, then clusters 3 and 4 and then cluster 1 is similar to 3 and so on.

6.5 Error

Cluster number	Number of hotspots	Percentage hotspots	Average distance	Unexplained variance	Percentage unexplained variance	Percentage grid cells	Grids cells per hotspot
1	1	0.2	0.0	0.0	0.0	0.1	2.0
2	35	8.2	0.5	15.8	3.7	7.4	4.9
3	2	0.5	0.5	0.4	0.1	0.2	2.0
4	40	9.4	0.6	17.7	4.1	6.9	4.0
5	48	11.2	0.4	23.9	5.6	15.4	7.4
6	36	8.4	0.4	16.4	3.9	9.3	5.9
7	49	11.5	0.9	23.5	5.5	6.2	2.9
8	45	10.5	0.4	27.1	6.4	16.5	8.4
9	4	0.9	2.9	17.6	4.1	1.4	8.0
10	9	2.1	0.2	11.1	2.6	15.5	39.3
11	33	7.7	0.9	14.7	3.4	3.7	2.5
12	54	12.7	1.0	25.8	6.0	6.0	2.6
13	34	8.0	1.2	29.7	7.0	5.8	3.9
14	35	8.2	1.4	32.2	7.5	5.4	3.5
15	2	0.5	34.5	25.8	6.0	0.2	2.0

Table 6.9: *General diagnostics of the clustering process*

In Table 6.9 the cluster number, number of hotspots and percentage of hotspots in this table are quite self explanatory. The average distance in this table refers to how uniform the cluster is, for example, the smaller the distance the more homogenous the cluster. Therefore from the data, cluster 8 is the most uniform and the hotspots within it therefore exhibit strong similarity. This is unlike cluster 15 which has a very heterogeneous appearance. The final two columns (percentage grids and grid cells per hotspot) are good indicators of the cluster form. Cluster 10 in this case has a high number of grid cells per hotspot. The percentage volume is the percentage of grid cells in each cluster. This indicates that clusters 5, 8 and 10 have a high number of grid cells, indicating that the clusters are spatially more extensive than the others. The variance in this table refers to the average distance times the percentage of records as the average distance is a good indicator of the homogeneity of the cluster as mentioned earlier. For example clusters with high variance indicators are clusters 14, 15 and 13. Although the average distance measure within cluster 15 is high. In this case the average distance will be used in order to determine a given cluster's homogeneity.

The final section looks at the way diagrams can be used to show the results from the clustering process. These include both a dendrogram and a minimum spanning tree. The spanning tree for road collision clustering process can be seen in Figures 6.4 and 6.5. A minimum spanning tree in this context is used to determine how each cluster links with each other. It is based on the fusion of clusters when the clustering algorithm is running to determine the grouping of clusters. It determines therefore which clusters are more alike than others. Table 6.8 shows this fusion process. Whereby, for example clusters 1 (A1) and 2 (A2) are fused together, then 3 (A3) and 4 (A4). Then 5 (A5) and 6 (A6) are fused together. These first four stages depict the most similar clusters for Group A. The process continues as one can identify from the table the next stage involves the fusing of Group D. The diagram below shows how this fusion works (Figure 6.4).

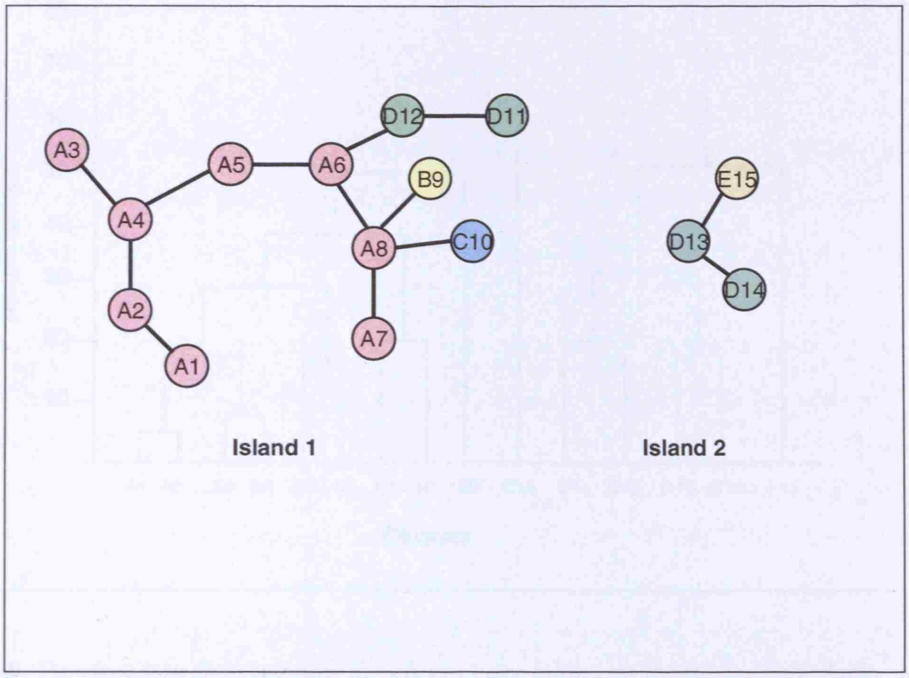


Figure 6.4: *Minimum Spanning Tree*

The final diagram (Figure 6.5) in this section is the dendrogram which shows the sequence in which groups are created from the clusters. In hierarchical clustering, dendrograms are used to visualise how the clusters are formed. We can see that this derives also from the fusion process; however the dendrogram is a different way of expressing the outcomes. The diagram below (Figure 6.5) shows how clusters 1 and 2 join and then 3 and 4 join together and these two joins form a larger link. This diagram shows that although clusters 9 and 10 formed separate groups they are most similar to Group A. This view allows the user to see how close one cluster is to another. For example clusters 1 and 2 are closely related, as are 3 and 4 and then 5 and 6 and 7 and 8. These closely related clusters then form groups and these 1-8 clusters form Group A. Clusters 9 and 10 are unique and form there own Groups as they are not closely related to any other cluster. Clusters 11 and 12 and then 13 and 14 are related to each other and therefore these form Group D. Finally cluster 15 is unique and stand alone and forms Group E.

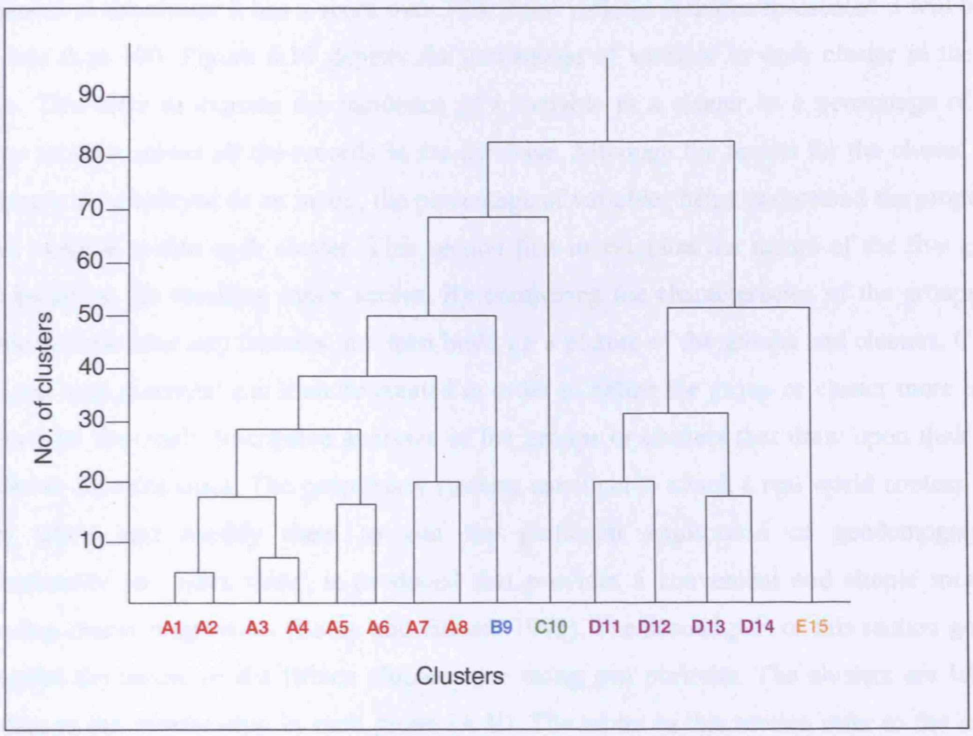


Figure 6.5: *Dendrogram showing how the clusters are joined on the basis of similarity*

6.6 Road Traffic Collision Group/Cluster Description

This section of the chapter explores the results of the clustering methodology and provides an interpretation of the cluster types and groups. Index scores are used to demonstrate the variables' representation in the cluster. Index scores have an average of 100; if the variable is over represented in the cluster it has a score over 100. If the variable is underrepresented it will have a score less than 100. Figure 6.10 depicts the percentage of variable in each cluster in the next section. This aims to express the incidence of a variable in a cluster as a percentage of their average incident across all the records in the database. Although the results for the cluster types and groups are displayed as an index, the percentage of variables helps understand the proportion of each variable within each cluster. This section first investigates the nature of the five groups (A-E) based on the resulting index scores. By comparing the characteristics of the groups it is possible to determine key features and then build up a picture of the groups and clusters. Cluster labels and 'pen portraits' can then be created in order to define the group or cluster more easily. Pen portraits are small descriptive analyses of the groups or clusters that draw upon their main identifiable characteristics. The proprietary systems use them to attach a real world context to the cluster labels and modify them to suit the particular application of geodemographics. Conventionally an 'index table' is produced that provides a convenient and simple means of comparing cluster diagnostics (Batey and Brown, 1995). The second part of this section goes on to describe the nature of the fifteen clusters also using pen portraits. The clusters are labelled according to the membership in each group (A-E). The tables in this section refer to the highest and lowest scoring indexes for the clusters. The complete clustering results can be found in the Appendix.

6.6.1 Group description

The groups created from the clustering process vary in size considerable, with groups A1 and A4 containing the majority of the clusters. This section endeavours to describe and explain the group indexes. The major difference between the groups and the clusters within them is the variance of the constituent variables. The variance for the variables in each group is considerably lower, because of the hierarchical structure of the clustering results. Therefore the description of the pen portraits is much less conclusive compared to the clusters.

Group A – 'Central London Pedestrians'

The highest index for Group A is pedestrians (index score = 127). This is the fundamental theme within this group as well the number of collisions involving a single vehicle, as the index scores are very small. There are few variables which are overtly over represented within this group. This group is depicted by the high propensity of collisions in urban centres where traffic is not travelling very fast, but where the pedestrians and traffic are not segregated.

Group B – 'High density vehicle damage'

This group is characterised by the over represented number of collisions in the cell indicating a high spatial density of collisions in these clusters. In relation to this there a high proportion of collisions in this group occur at the weekend, particularly on a Saturday when it is dark, possibly early morning or late evening. The results suggest careless driving in conditions when there is not a lot of traffic and thus cars are more likely to cause damage when the driver is not concentrating.

Group C 'Cyclists in danger'

Cyclists are vastly over represented in this group, having an index score of 247, and an equally strong variance of 70. In conjunction with this there is also an over representation of collisions involving one vehicle that result in a fatal injury.

Group D 'Multiple main road collisions'

This group has a very different composition to the others in so far as it is characterised by a high number of casualties and vehicles, with collisions that occur between 4-7am being slightly over represented. They are also characterised by a high proportion of collisions occurring on major roads.

Group E 'Weekend risk takers'

This group has very high index scores, especially for collisions resulting in three casualties, which has a very strong variance score as well. Coupled with three casualties there is an over representation of the average number of vehicles in the collisions. In summary it would be feasible to conclude that although these collisions have an above average number of vehicles and casualties there are not as many as in the previous group. This group is characterised by the over representation of pedestrians, and high number of vehicles and casualties, as well as collisions being more likely to occur on a Sunday than any other day of the week.

6.6.2 Cluster description

6.6.2.1 Bar charts for index values of cluster variables

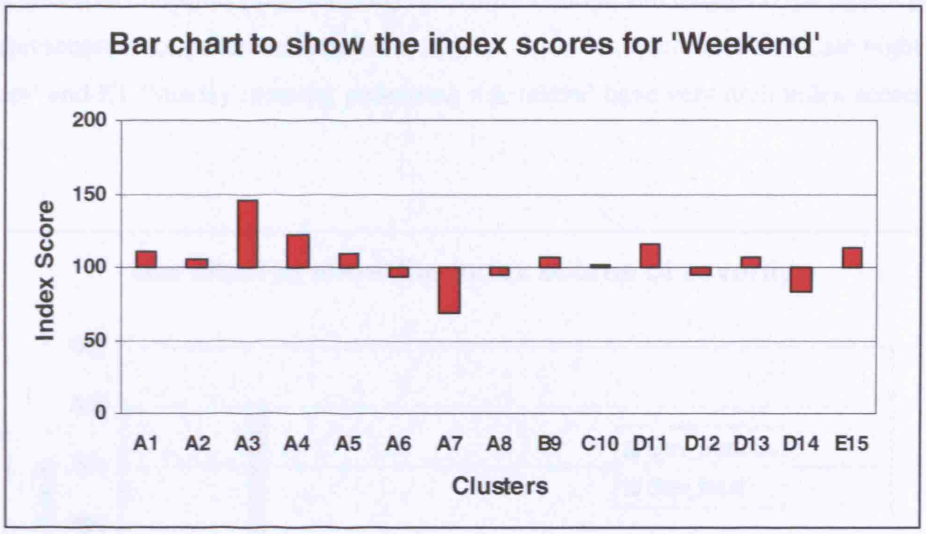


Figure 6.6: *Table to show index scores for the variable 'Weekend' by cluster*

This chart (Figure 6.6) shows the differences in index scores for the variable 'weekend' within the 15 clusters. Clearly the cluster A3 'Saturday Morning Leisure Hotspots' has the highest index score for this variable at 145. Notably Cluster A7 'Weekday rush hour pedestrian hotspots' has a strong under representation of this variable which has an index score of 69.

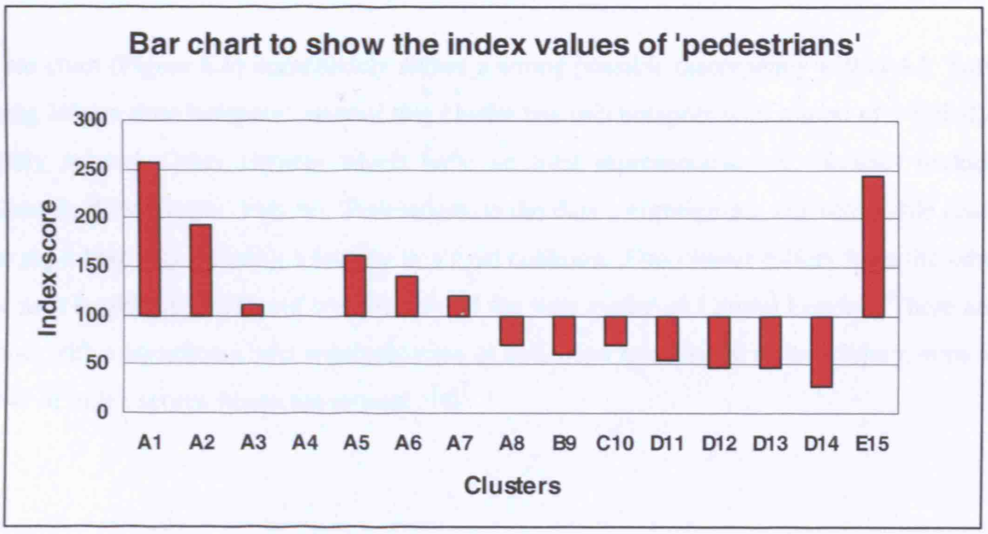


Figure 6.7: *Table to show index scores for the variable 'Pedestrian' by cluster*

This bar chart (Figure 6.7) clearly shows the high propensity of pedestrian related collisions in Group A which has hotspots predominantly in Central London and Groups B –D have a pattern of under representation of pedestrian related collisions. Specifically clusters A1 'Late night football supporters' and E1 'Sunday morning pedestrian risk takers' have very high index scores for this variable.

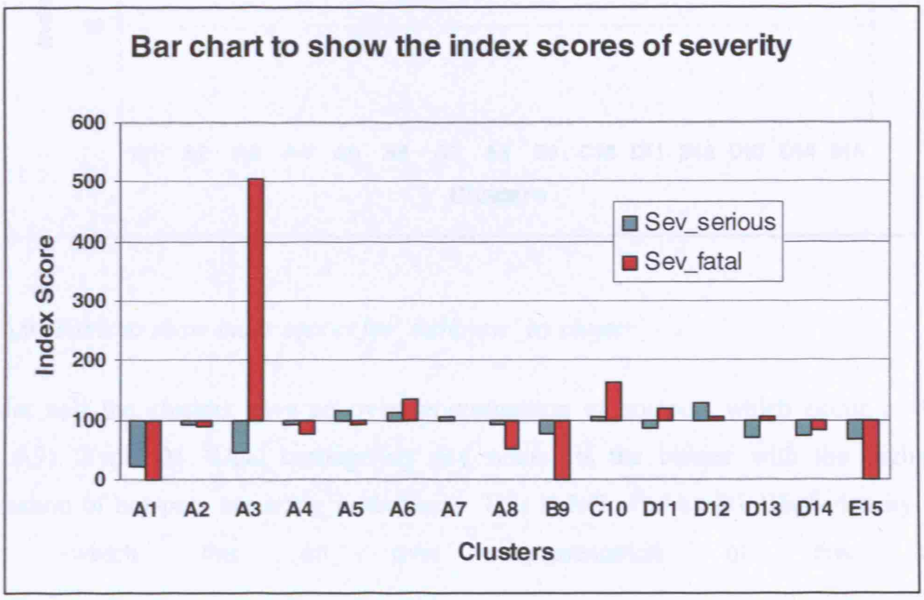


Figure 6.8: *Table to show index scores for severity 'Killed or seriously injured' by cluster*

This bar chart (Figure 6.8) immediately shows a strong possible discrepancy within A3 'Saturday morning leisure time hotspots', overall this cluster has two hotspots with a total of 2 killed and 4 seriously injured. Other clusters which have an over representation of fatalities include C1 'Cyclists in Westminster' and A6 'Pedestrians in the dark', highlighting the vulnerable road user as having a high risk of being a fatality in a road collision. This cluster differs from the others in so far as it is place specific and centred around the very centre of Central London. There are few hotspots with a significant over representation of collisions resulting in serious injury, with a high number of index scores which are around 100.

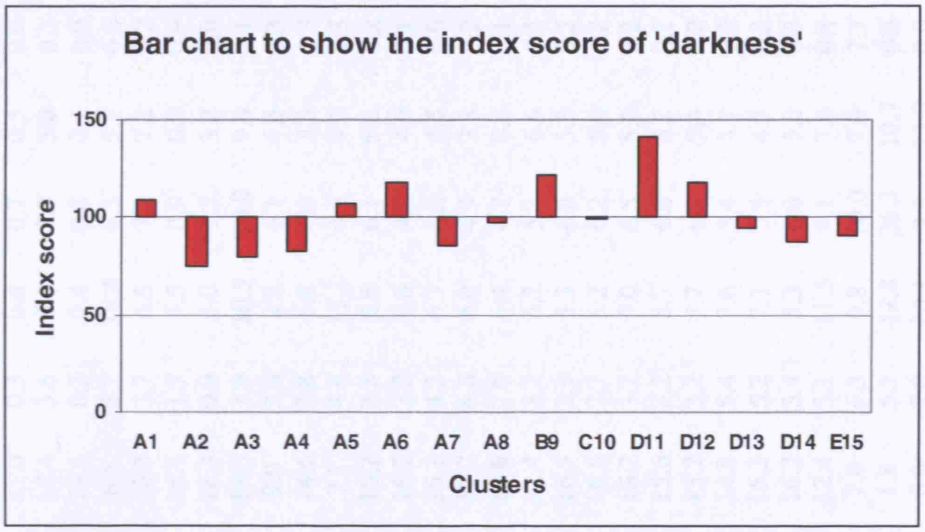


Figure 6.9: *Table to show index scores for 'darkness' by cluster*

Just under half the clusters have an over representation of hotspots which occur in darkness (Figure 6.9). Type D1 'Dual carriageway joy riders' is the cluster with the highest over representation of hotspots occurring in darkness. This is followed by B1 'High density careless drivers', which has an over representation of this variable.

Variable	A1	A2	A3	A4	A5	A6	A7	A8	B9	C10	D11	D12	D13	D14	E15
Number of collisions	0.0	1.2	0.0	1.3	4.4	3.4	0.5	6.1	0.2	81.0	0.3	0.4	0.7	0.5	0.0
Collisions per cell	0.1	7.5	0.2	6.5	15.2	8.8	6.3	16.1	4.9	14.4	3.6	5.5	5.7	5.0	0.2
Number of cells	0.0	1.2	0.0	1.3	4.6	3.5	0.6	6.5	0.5	79.8	0.3	0.4	0.8	0.5	0.0
Number of speed cameras	0.0	4.4	0.0	5.1	12.9	12.8	7.0	13.3	1.3	6.5	5.7	18.4	6.3	5.7	0.6
Road length	0.1	6.4	0.2	6.2	13.5	8.6	6.4	16.3	1.2	16.3	3.7	6.5	7.2	7.2	0.2
Number of schools	0.0	9.6	0.0	12.2	15.6	7.8	10.3	19.9	3.5	14.1	1.7	3.5	0.9	0.0	0.9
Tube stations	0.0	12.0	0.0	9.1	21.2	12.0	6.0	14.0	0.0	18.3	0.0	3.0	1.5	3.0	0.0
Cycle lane length	0.0	6.1	0.1	6.4	12.1	5.7	7.6	13.7	0.6	14.5	3.0	10.5	10.0	9.6	0.1
Traffic lights	0.0	3.5	0.3	6.8	15.7	11.2	9.3	19.2	2.4	9.0	3.0	8.8	6.1	4.5	0.3
Bus stops	0.2	11.8	0.2	7.4	18.7	8.7	5.7	13.9	0.7	14.6	3.8	6.4	4.3	3.5	0.3
Pedestrian crossings	0.0	4.4	0.5	6.3	16.2	12.1	9.6	17.8	1.4	7.1	4.4	11.1	4.5	4.3	0.3
Slight	0.1	7.5	0.2	7.1	15.1	9.2	6.2	16.7	1.1	15.2	3.8	5.8	6.1	5.6	0.2
Serious	0.0	7.0	0.1	6.4	17.6	10.6	6.6	15.5	1.1	16.2	3.2	7.6	4.0	4.0	0.1
Fatal	0.0	6.6	0.9	5.3	14.0	12.4	6.2	8.4	0.0	25.1	4.3	6.3	6.2	4.5	0.0
Time 0000 - 0359	0.0	4.7	0.3	5.7	18.1	11.4	5.6	13.9	1.6	17.7	6.8	6.0	4.9	3.3	0.2
Time 0400 - 0659	0.0	4.6	0.1	4.8	13.8	10.9	5.3	19.0	2.2	12.6	7.0	6.4	4.7	8.6	0.1
Time 0700 - 0959	0.1	6.7	0.2	7.0	13.9	10.2	7.4	18.8	1.5	12.4	3.2	6.7	5.2	6.6	0.3
Time 1000 - 1159	0.1	9.8	0.3	7.7	14.1	8.6	5.9	16.5	1.2	16.3	2.9	5.3	6.0	5.3	0.2
Time 1200 - 1459	0.1	9.1	0.2	8.0	16.0	7.9	5.9	16.0	1.1	16.6	2.7	5.2	6.2	5.0	0.2
Time 1500 - 1659	0.1	8.9	0.1	8.1	15.3	8.4	6.7	14.6	1.4	16.2	3.2	5.0	6.5	5.3	0.2
Time 1700 - 1859	0.1	7.2	0.2	6.7	15.4	8.2	7.1	17.1	1.3	15.6	3.2	5.7	6.0	6.2	0.1
Time 1900 - 2059	0.1	6.7	0.2	6.0	15.6	10.8	5.4	15.8	1.7	15.2	3.2	7.7	6.2	5.4	0.2
Time 2100 - 2359	0.2	5.5	0.1	6.0	16.8	11.1	5.2	16.5	1.6	14.8	5.4	7.6	5.4	3.7	0.2
Dark	0.1	5.6	0.1	5.8	16.5	11.0	5.4	16.0	1.7	15.3	5.2	7.1	5.5	4.7	0.2
Casualties = 1	0.1	7.6	0.2	7.0	16.1	9.5	6.4	16.1	0.5	16.2	3.4	5.3	6.0	5.0	0.5
Casualties = 2	0.0	6.4	0.1	6.5	12.0	8.9	5.8	13.9	0.2	12.4	5.1	11.5	9.1	7.6	0.4
Casualties = 3	0.0	5.3	0.1	6.0	10.0	7.5	6.0	15.8	0.1	7.4	6.3	9.8	10.0	7.9	7.7
Casualties = 4	0.0	4.0	0.0	6.8	13.8	3.7	4.3	10.0	0.0	1.8	5.1	12.8	26.3	10.7	0.8
Casualties = 5	0.0	6.0	0.0	4.3	13.3	12.8	1.8	12.0	0.0	0.0	7.0	20.1	9.8	13.0	0.0
Casualties = 6	0.0	6.1	0.0	0.0	29.3	3.7	6.1	11.0	0.0	0.0	11.6	9.8	19.5	3.0	0.0
Casualties per cell	0.1	7.2	0.2	6.9	14.9	9.2	6.2	15.6	0.4	14.7	4.0	6.8	7.2	5.8	0.9

Variable	A1	A2	A3	A4	A5	A6	A7	A8	B9	C10	D11	D12	D13	D14	E15
Monday	0.1	7.6	0.2	6.4	15.8	9.9	5.8	15.3	1.1	13.2	4.6	7.8	6.3	5.8	0.2
Tuesday	0.1	7.3	0.1	6.8	14.3	9.3	7.4	16.1	1.5	15.1	3.5	6.4	6.0	6.0	0.2
Wednesday	0.1	7.5	0.2	6.0	14.9	9.4	6.5	18.7	1.5	15.2	3.3	5.8	5.4	5.6	0.1
Thursday	0.1	6.8	0.1	6.1	14.9	9.8	7.6	16.7	1.1	17.0	3.1	5.8	5.2	5.5	0.2
Friday	0.0	7.2	0.1	6.2	14.2	9.4	7.5	17.6	1.5	16.4	3.0	5.4	5.4	5.8	0.2
Saturday	0.1	7.7	0.3	7.5	16.1	9.5	4.5	16.5	1.8	16.3	3.3	5.9	5.9	4.5	0.1
Sunday	0.1	8.0	0.2	9.6	17.5	7.9	4.1	15.0	1.1	13.9	5.3	5.6	6.7	4.6	0.3
Weekend	0.1	7.8	0.3	8.4	16.7	8.7	4.3	15.7	1.5	15.6	4.2	5.7	6.2	4.5	0.2
Vehicles = 1	0.2	10.6	0.2	6.9	18.6	9.6	6.0	12.7	0.3	23.0	2.2	2.9	4.2	2.4	0.3
Vehicles = 2	0.0	5.7	0.2	6.9	13.9	9.4	6.6	18.0	0.5	11.6	4.5	7.8	7.7	6.6	0.7
Vehicles = 3	0.1	5.5	0.1	6.0	12.8	7.9	5.1	15.8	1.1	9.5	5.5	7.9	8.1	13.9	0.6
Vehicles = 4	0.0	4.3	0.0	7.0	9.6	4.5	3.9	14.6	0.0	0.6	2.0	12.0	12.3	28.3	0.9
Vehicles = 5	0.0	5.6	0.0	3.9	2.2	0.0	4.4	12.2	0.0	0.0	3.3	26.1	7.8	34.4	0.0
Average number vehicles	0.1	6.7	0.2	6.8	14.7	9.3	6.3	16.6	0.5	13.7	4.1	6.9	7.0	6.7	0.6
Fine no winds	0.1	7.7	0.2	7.0	15.6	9.1	6.4	16.1	0.4	15.4	3.6	6.0	6.6	5.1	0.6
Fine with winds	0.2	6.3	0.1	6.9	15.5	9.9	5.8	15.3	0.6	15.4	4.4	6.6	5.7	7.1	0.4
Raining no winds	0.0	6.5	0.0	7.9	13.0	8.6	8.6	13.0	0.0	0.0	0.0	23.0	2.2	17.3	0.0
Raining with winds	0.0	4.9	0.0	4.6	16.7	10.5	4.6	22.3	0.0	4.0	3.4	8.1	13.0	7.8	0.0
Snow	0.0	3.7	0.0	5.0	10.3	13.4	13.9	16.1	0.0	0.8	5.5	3.9	10.8	16.6	0.0
Other weather	0.0	5.7	0.5	6.5	13.9	4.7	3.8	10.4	0.5	27.5	8.2	5.7	7.8	4.3	0.4
Unknown weather	0.0	5.1	0.9	3.9	10.9	23.4	8.9	9.5	1.4	18.6	1.7	6.5	3.8	4.9	0.4
Pedestrians	0.2	14.2	0.2	6.9	24.9	13.1	7.6	11.7	0.9	10.9	2.0	2.8	2.7	1.4	0.4
Cyclists	0.0	3.9	0.1	6.6	10.1	7.4	5.5	19.6	0.6	38.1	1.7	2.7	1.5	2.0	0.1

Table 6.10: *To show the percentage of variable in each cluster*

The table above (Table 6.10) depicts the percentage of variable within each cluster. For example cluster A1 has a very small proportion of the variable percentage this is due to the small number of hotspots within the cluster. Therefore the bigger the cluster (i.e. the number of hotspots) the more likely it is to have higher percentage of the variance. This table provides a useful interpretation of the variables and their loading across the cluster types. Cluster A5 has the highest proportion of serious collisions. This cluster is characterised by pedestrian collisions within zone 1, and this information assists with understanding the proportion of serious collisions with regards to the other clusters. Similarly cluster type C10 accounts for 25% of the total fatal collisions. This cluster is characterised by the large spatial extent of its constituent hotspots that together cover a large area of Central London (Oxford Street, Tottenham Court Road, Soho, Charing Cross Road and Edgware Road). Cluster C10 has 38% of the collisions involving cyclists. This is a large proportion which is not indicated in the index scores. Overall, it would indicate therefore that the combined use of both index scores and percentage of the variable within each cluster (Table 6.10) creates a useful and accurate interpretation of the cluster.

6.6.2 Characteristics of the clusters

A1 – 'Late night football supporters'

This cluster only consists of one hotspot, approximately 0.23% of the total number of records. The data for this cluster suggests there is a universal uniformity and no error. Out of all the variables 37% had an index score of zero, including variables such as cyclists, weekdays excluding Wednesday and Sunday, early and late afternoon and morning. Table 6.11 shows the seven highest and lowest index scores for A1 and their associated variance. The pattern of collisions in this type involves pedestrians and single vehicles late at night. Other over represented attributes included collisions being serious in severity, involving one casualty and occurring on a Saturday.

HIGH			LOW		
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Pedestrians	258	62.11	Snowing	0	3.5
Bus stops	216	5.65	Raining with wind	0	7.77
Vehicle = 1	194	57.42	Casualties = 5	0	8.25
Raining no high winds	191	10.42	Fine w' wind	0	6.21
Time 2100 - 2359	189	30.54	Severity = fatal	0	9.16
Wednesday	152	9.37	Casualties = 6	0	4.72
Thursday	132	16.19	Tube station	0	4.03

Table 6.11: *Characteristics of Cluster Type A1*

This hotspot is located in Newham close to the Upton Park football stadium at a junction where the B165 meets the B167. The map (Figure 6.10) shows the exact location of the hotspot. There is a higher than average number of collisions on a Wednesday and Saturday possibly linked to football matches and supporters. It would indicate a high risk time of late at night possibly when the pubs have closed and pedestrians are at risk due to being intoxicated and poor visibility. Although in the past there have been campaigns to reduce the amount of hooliganism of football supporters there has been at best only limited attempt to target the safety of football supporters especially in Outer London where the roads tend to have higher speed limits.

Figure 6.10: *A1 – 'Late night football supporters'*

A2 – 'Inner city pedestrian risk takers'

This cluster consists of 35 hotspots, approximately 8.2% of the total number. The volume per record which is the average number of cells per hotspot is 4.8 which is a little over average and indicating an average hotspot size of five cells. The average distance of this cluster is 0.49, indicated a predominantly homogenous cluster. In this cluster there are a high proportion of environmental variables which contribute to the pattern of collisions. The strongest pattern in this cluster is the high proportion of pedestrians and the number of vehicles (1) and the time of day which is approximately between 10.00 and 15.00 (Table 6.12). This cluster can be described as having a high proportion of pedestrian collisions involving one vehicle close to tube stations, bus stops and schools during the day, indicating the causes to be an inner city London phenomenon. From the original count data it is possible to assess the reliability of the variables with low variance. The count data show that this cluster accounts for 12% of the total number of hotspots near tube stations (an average is 6% for any variable as this is the total divided by the number of clusters which is 15). The high propensity of tube stations suggests it is an inner city phenomenon as there are more tube stations in Inner London. Coupled with this the proportion of schools in this cluster from the total is 12%.

HIGH			LOW		
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Pedestrians	192	62.11	Pedestrian	59	9.7
Tube stations	161	4.03	Cyclists	53	75.69
Bus stops	159	5.65	Casualties = 4	53	6.95
Vehicle = 1	143	57.42	Rain w' wind	50	7.77
Time 1000 -1159	132	12.14	Traffic lights	47	7.44
Schools	129	4.37	Number cells	16	83.02
Time 1200_1459	122	24.33	Number	16	82.83

Table 6.12: *Characteristics of Cluster Type A2*

Spatially there are a large proportion of hotspots in Lewisham and Catford. These two areas are fairly highly populated with pedestrians as they serve as shopping centres for areas of the South of London. In conjunction with this there are a number of hotspots along Kensington High Street (Figure 6.11), Chelsea and Hammersmith all of which occur of A-roads. The locations of these hotspots seems to be clustered within Central London, however they are found out as far as Streatham, Wembley, Willesden and Tottenham and they are very centrally located within all these areas. This suggests it is associated with busy urban shopping centres that are in the localised 'town centres' of London.

Figure 6.11: *A2 'Inner city pedestrian risk takers' (Kensington and Chelsea)*

A3 – 'Saturday morning leisure time hotspots'

Type A3 is similar to A1 insofar as it only has two hotspots (0.47% of the total records). It has on average 2 cells per hotspot. The error is negligible considering the size of the cluster size. This cluster displays some unique characteristics insofar as the index scores are predominantly high for all the over represented attributes and in association the variance is relatively low. In summary this cluster has a predominance of fatal collisions at pedestrian crossings, in 'unknown' or 'other' weather, with a high occurrence of these on a Saturday morning between 10 -12pm (Table 6.13). Other over represented variables in this cluster include Sundays and collisions occurring between 5-7pm. The high proportion of collisions occurring in unknown or other weather suggests that the weather was not reported at the scene of the collision and therefore leads to a high inaccuracy of the nature of the weather in this cluster.

	HIGH			**LOW**	
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Unknown weather	540	14.57	Casualties = 5	0	8.24
Severity = Fatal	504	9.16	Vehicles = 5	0	12.57
Pedestrian	274	9.7	Fine with wind	0	6.21
Other	263	12.12	Speed cameras	0	7.02
Saturday	167	18.32	Vehicles_ 4	0	5.2

Traffic Lights	163	7.44	Casualties = 4	0	6.95
Time 1000 - 1159	150	12.14	Raining with wind	0	7.77

Table 6.13: *Characteristics of Cluster Type A3*

Looking more closely at the two locations of these hotspots indicates pattern. The first of the two hotspots is in Brentford (see Figure 6.12) which is located at a junction close to a leisure centre. Therefore it is feasible to suggest that this type is associated with the people using the leisure centre, with particular increase on a Saturdays when many children and adults would visit a leisure centre. The second hotspot (Figure 6.13) is in Kingston and is also located on a junction close to a leisure centre and a museum. This collision hotspot has different dynamics in so far as there are no fatal injuries at this site. Overall, this cluster would indicate a strong pattern of collisions related to the surrounding land use and the number of pedestrians at risk from being hit by a vehicle resulting in serious or fatal injury.

Figure 6.12: *A3 'Saturday morning leisure time hotspots' (Brentford)*

Figure 6.13 *A3 'Saturday morning leisure time hotspots' (Kingston-Upon-Thames)*

A4 – 'Sunday afternoon car drivers'

Type A4 has 40 records (9%) of the total number of records. It has a volume per record of 4, indicating an average cluster size of 4 cells which is fairly heterogeneous, and an average distance of 0.6, again indicating that this cluster is fairly uniform in character. The strongest attribute of the cluster is the over representation of hotspots which have collisions occurring on a Sunday afternoons and a slight over representation of collisions on a Saturday. The count data reveal there to be a high proportion of hotspots which have schools in the vicinity, however because these collisions predominantly occur on at the weekend, it is unlikely that school children would be over represented within this cluster. The count data reveals this cluster has 18.6% of the total 'Sunday' variable and 17.2% of the Saturday variable. This is an extremely high proportion and supports the findings of the index scores for this variable. The proportion of tube stations was 15% which could indicate that this cluster has a central London locational pattern (Table 6.14). It is important to recognise that type A4 differs from D3 with respect to location and number of casualties.

HIGH	**LOW**

Variable	Index	Variance	Variable	Index	Variance
Schools	175	4.37	Fine w' wind	67	6.21
Sunday	139	40.7	Casualties = 5	61	8.24
Tube stations	130	4.03	Vehicles = 5	56	12.57
Weekend	122	49.95	Unknown	56	14.57
Time 1500 –	117	12.13	Number	19	83.02
Time 1200 –	115	24.33	Number cells	19	82.83
Snowing	114	3.5	Casualties = 6	0	4.72

Table 6.14: *Characteristics of Cluster Type A4*

Spatially type A4 covers a wide area within London, with the majority of hotspots occurring in Ealing (4) and Lewisham (4), both lying south and West of Central London. A predominant hotspot lies at a large junction in Lewisham. Generally this type appears to be an Outer London phenomenon indicated by the average representation of pedestrians and under representation of cyclists. There are only three hotspots in Central London and these occur at Paddington, Shoreditch and Islington. In Ealing there appears to be recurring hotspots along the A4020 (see Figure 6.14). Overall the hotspots in this type tend to be route hotspots, on arterial roads coming into and out of London.

Figure 6.14: *A4 'Sunday afternoon drivers in Central London'*

A5 – 'Illicit late night Zone 1 pedestrians'

Initial appearances shows type A5 appears similar to type A1, insofar as there is a high index score of collisions involving pedestrians and one vehicle. This situation is not uncommon in a city such as London with the high number of pedestrians and generally the number of collisions in London has a high proportion of collisions involving pedestrians and single vehicles. In this cluster however, there are an over- represented number of collisions involving 6 casualties. This cluster consists of 48 of the total number of records (428). Its volume per record is relatively large, at 7.3 indicating a seven cell average for each hotspot. The average distance is very low at 0.36 indicating all the hotspots are of a similar nature. The count data show that type A5 only accounts for 8% for the total number of collisions involving 6 casualties, which is only a small percentage higher than the average. Therefore type A5 has a higher than average number of collisions occurring between midnight and 3am early on a Sunday morning. Other over represented attributes of this cluster include collisions resulting in serious injury. The count data show that A5 accounts for 12% of the total collisions at pedestrian crossings, a significant proportion of the total count data, in addition to which this cluster accounts for 12% of all serious injuries, again a significant proportion of the total count data, above the average of 7%. This might suggest pedestrians (possibly near bus stops, 9% of total number of bus stops) being hit early morning, after being out socialising on Saturday night with the possibility of drinking and taking drugs. In summary type A5 is associated with pedestrians trying to cross busy roads by illicit means late at night (Table 6.15).

HIGH			LOW		
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Casualties = 6	190	4.72	Rain with wind	67	7.77
Pedestrians	161	62.11	Cyclist	66	75.69
Tube stations	138	4.03	Casualties = 3	65	71.81
Bus stops	121	5.65	Vehicles = 4	62	5.2
Vehicles = 1	121	57.42	Number of	30	12.57
Time 0000 – 0359	117	33.83	Number of	28	83.02
Sunday	114	40.7	Vehicles = 5	14	82.83

Table 6.15: *Characteristics of Cluster Type A5*

The London boroughs with high proportions of A5 include Islington (5), Southwark (5), Camden (4), Haringey (4) and Kensington (4). This shows a strong Central London distribution, which supports the idea of pedestrians who have been drinking in Central London and then try to cross busy roads in order to get home or to a bus stop. There is a prominent Camden hotspot associated with high numbers of bars and clubs in this area (see Figure 6.15). There is also a large hotspot at a junction between Hammersmith and Kensington.

Figure 6.15: *A5 'Illicit late night Zone 1 pedestrians (Camden)'*

A6 – 'Pedestrians in the dark'

This cluster has 36 records (8.43%) of the total. The average distance is 0.4 which means that the cluster is homogenous. The volume per record is 6 which means that the average hotspot size in this cluster is 6 cells which are fairly large in comparison to the rest of the types. A potential drawback of this cluster and its associated attributes lies with the fact that the variances of all the high index score variables are very low apart from the involvement of pedestrians in the collision. There is a high index score for collisions occurring in unknown weather and this index score drops quite considerably for the next highest index score, 'raining with high winds', which has an index score of 144. Although the variance explained for the number of collisions occurring in rain with high winds is low it should not be discredited in the explanation of the cluster, as with the other variance attributes. Other over represented variables includes Time 0000-0359 and Time 2100-2359 and collisions resulting in serious injury. From the count data this cluster has 13% of the total number of speed cameras indicating a higher than average proportion of speed cameras and therefore a higher propensity for hotspots to occur on main roads as it is on these roads that speed cameras are predominantly sited. The count data also revealed that 12.5% of hotspots in this cluster occurred at or near pedestrian crossings indicating a strong validity of the index scores. In conjunction with this there was also an over representation within the count data of fatal hotspots. Therefore, in summary the types of collisions involved in type A6 are more likely to be fatal and occur near pedestrian crossings and near speed cameras and therefore on main roads (Table 6.16). With regard to times of day, collisions are more likely to occur between midnight and 4am (index score = 122).

HIGH			LOW		
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Unknown weather	250	14.57	Other	51	12.12
Raining with high	144	7.77	Vehicle = 4	49	5.2
Pedestrians	141	62.11	Casualties =	40	6.95
Casualties = 5	137	8.24	Casualties =	39	4.72
Number of speed	137	7.02	Number cells	38	82.83
Severity = fatal	133	9.16	Number	36	83.02
Pedestrian	130	9.7	Vehicles = 5	0	12.57

Table 6.16: *Characteristics of Cluster Type A6*

The Borough with the largest number of hotspots in A6 is Lambeth (7) followed by Camden (4), Hammersmith (4) and Tower Hamlets (4) indicating a strong Central London pattern. This cluster covers the Euston Road from the top of Tottenham Court road towards and beyond Kings Cross. The Euston road near to Kings Cross is a very busy area with increased number of vehicles and pedestrians interacting in a very dynamic environment There are many commuters and students using the road and although Euston road is well lit with pedestrian crossings the propensity of these types of collision suggests that the lighting needs to be addressed coupled with the use of pedestrian crossings.. These hotspots are predominately route hotspots, with other hotspots in Lambeth and Clapham (Figure 6.16).

Figure 6.16: *A6 'Pedestrians in the dark'*

A7 – 'Weekday rush hour pedestrian hotspots'

This cluster constitutes 49 of the total number of records (11.6%), with an average distance of 0.9, making the cluster heterogeneous in nature. The error is high, with a score of 23.5 (5.5%), this is the percentage of unexplained variance within the cluster. However the volume per record is 2.9 making the average size of each hotspot in the cluster quite small. As with the previous cluster, A7 has a high proportion of index scores with fairly low variance. However, from these results a typology can be produced. The most important pattern is the high number of collisions occurring on Thursday and Friday (Index score 120), with a variance which is significantly higher than the other attributes, indicating that this pattern is strong. Other high index scoring variables include T 0700 – 0959 and T 1500 – 1659 suggesting a rush hour pattern. This time also represents when school children leave school and spend time walking home and interacting in the road environment. The cluster has a high proportion of collisions occurring near schools and at pedestrian crossings possibly with traffic lights. The count data reveal that this cluster contains 10% of all hotspots near to schools, which compared to the average (100/15 – the total number of clusters) which is 7%; this is a high proportion of hotspots close to schools. Likewise this cluster consists of 9.8% hotspots at or near pedestrian crossings, and 9.6% at traffic lights. One of the reasons for the high proportion of pedestrian crossings and traffic lights is because this cluster usually occurs in isolated areas (non route) involving pedestrians at peak rush hour times. This cluster has an under representation of cyclists involved in the hotspots, indicating a stronger pedestrian risk of being involved in a collision in these areas (Table 6.17).

HIGH			LOW		
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Raining with	223	7.77	Casualties = 4	68	6.95
Schools	165	4.37	Sunday	66	40.7
Pedestrian	153	9.7	Vehicles = 4	62	5.2
Traffic lights	148	7.44	Other weather	60	12.12
Unknown	142	14.57	Casualties = 5	28	8.24
Snowing	138	3.4	Number cells	9	82.83
Thursday	122	16.19	Number	9	83.02

Table 6.17*: Characteristics of Cluster Type A7*

Figure 6.17 shows that this cluster dominants Central London, especially the areas of Westminster and London Bridge which are high commuter areas in Central London. With regard to borough propensities, Lambeth has the highest number of A7 clusters (7), followed by Kensington and Chelsea (6) and Tower Hamlets (5). There are a few isolated hotspots in Hampstead, Camden and Islington, however the majority of the locations within Central London in prominent commuter areas.

Figure 6.17: *A7 'Weekday rush hour pedestrian hotspots' (Westminster and London Bridge)*

A8 – 'Morning commuting cyclists at rush hour'

Similar to A7, this cluster has a large proportion of the records (45, 10.5%), with a small average distance confirming the strong heterogeneous nature of the cluster. The average number of grid cells per hotspot is 8 which is fairly large compared to the other clusters. This cluster has a high number of cyclists being involved in collisions in the early, peak travel time in the morning. The cyclist attribute has an index score of 119. The variance explained for this variable is very high at 75.69, which means we can rely on this attribute being largely accurate. This is in contrast to the under representation of pedestrians in this cluster. A similar pattern can be applied to the attributes of time of day, which although having an index score of 115 and 114 respectively also each have high variance scores (Table 6.18). This indicates that the majority of the hotspots occur in the morning rush hour to work, during the week (the weekend variable is under represented, with an index score of 95). The count data reveal that this cluster is accountable for 20% of all the hotspots near to schools which is high, and indicates that this cluster has a high propensity for school children riding to school and commuters cycling to work. Similarly the count data support the strong presence of cyclists, with this cluster accounting for 19.5% of all the hotspots with cyclists involved.

	HIGH			**LOW**	
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Fine with winds	135	6.21	Pedestrians	71	62.11
Schools	121	4.37	Casualties = 6	67	4.72
Cyclists	119	75.69	Other weather	63	12.12
Traffic lights	117	7.44	Casualties = 4	61	6.95
Time 0400 – 0659	115	25.52	Unknown	58	14.57
Time 0700 – 0959	114	21.75	Severity = fatal	51	9.16
Wednesday	114	9.37	Number	39	82.83

Table 6.18: *Characteristics of Cluster Type A8*

Spatially, the highest number of hotspots occur in Hackney (8), followed by Southwark (5) and Westminster (5). Generally this cluster is similar to A7 in terms of the predominant Central London pattern to the location of the hotspots. In comparison to A6, there is an A8 hotspot covering the A40 past Marylebone and Regent's Park (see Figure 6.18). There are also two other major hotspots in Bermondsey and Waterloo, reflecting the arterial road pattern of this cluster. There are approximately 10 major hotspots in Central London, becoming progressively smaller and more isolated with increased distance from Central London. Figure 6.18 shows this pattern in Hackney and Stoke Newington.

Figure 6.18: *A8 'Morning commuting cyclists at rush hour' (Southwark)*

B9 – 'High density careless weekend drivers'

B1 is a very small cluster with 4 records (1% of the total). It has an above average distance score indicating a level of heterogeneity within the cluster. The volume per record is 8 indicating a large average hotspot size for this cluster. Cluster B9 has a number of high index scoring attributes with corresponding high variance scores. This cluster has a high propensity of collisions occurring on a Saturday in the dark, particularly early morning. There are also a high proportion of collisions in this cluster which occur early evening (7pm – 9pm) possibly on a Saturday, and in the dark, indicating a winter phenomenon. There is a high index score (351) and a strong variance for collisions per cell indicating there is a high density of collisions occurring in each of the cells of the hotspots in the cluster. There is an under representation of cyclists and pedestrians, and although there is a high number of collisions per cell there is an under representation of casualties per cell indicating a high number of collisions not resulting in serious or fatal injury (Table 6.19).

	HIGH			LOW	
Variable	Index	Variance	Variable	Index	Variance
Collisions per	351	63.63	Tube stations	0	4.03
Schools	253	4.37	Snowing	0	3.5
Traffic Lights	172	7.44	Vehicles = 5	0	12.57
Time 0400 –	154	25.42	Casualties = 5	0	8.24
Saturday	125	18.32	Casualties = 6	0	4.72
Dark	122	55.66	Casualties = 4	0	6.95
Time 1900 -	120	11.56	Severity = fatal	0	9.16

Table 6.19: *Characteristics of Cluster Type B9*

The cluster has its largest hotspot along Upper Street in Angel (see Figure 6.19). There are hotspots in Wandsworth, Islington, Kensington and Hounslow, indicating both a Central and Outer London pattern. From the maps this is clearly a main road phenomenon.

Figure 6.19: *B9 'High Density careless weekend drivers' (Upper Street, Islington)*

C10 – 'Cyclists in Westminster'

The characteristics of this cluster are very unique. It has 10 records (only 2.1% of the total), and an average distance of 0.2 indicating particularly strong uniformity. The average number of grid cells per hotspot is 39.3 which is the highest out of all the clusters. This is largely due to the nature of the cluster covering a large proportion of Central London. This cluster, unlike all the others has a significantly high index score for both the number of collisions and number of cells. This indicates therefore that the hotspots which occur in this cluster have a high number of collisions and are spatially extensive. Many of these cluster 10 hotspots occur in Central London, particularly Westminster where there is a large traffic flow and the hotspots cover large areas. The cluster also has a high number of collisions involving only one vehicle, likely to result in a fatal injury. The over representation of tube stations, is because the majority of the hotspots in this cluster occur in Westminster and there are many tube stations in Westminster, so the count in the grid cells would be higher for this variable possibly when it would not impact the cause of the collisions. However, what is missed from the index scores is the unusually low propensity for cyclists involved. From the index scores, cyclists are under represented within this cluster (Table 6.20): however the count data reveal that this cluster account for 37% of all cyclists' collisions. It also consists of nearly 17% of all the fatal collisions, which is significantly higher than average.

HIGH			LOW		
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Number	524	83.02	Casualties = 4	11	6.95
Number cells	516	82.83	Rain with wind	5	7.77
Other weather	178	12.12	Vehicles = 4	4	5.2
Severity = Fatal	162	9.16	Snowing	0	3.5
Vehicles = 1	149	57.42	Vehicles = 5	0	12.57
Unknown	120	14.57	Casualties = 5	0	8.24
Tube stations	118	4.03	Casualties = 6	0	4.72

Table 6.20: *Characteristics of Cluster Type C10*

The reason for this cluster's uniqueness is because it consists of the largest hotspot covering Soho, Oxford Street, Edgware Road and Tottenham Court Road (see Figure 6.20). Surrounding this hotspot are several smaller ones in Hackney, Westminster, Paddington and Westbourne Grove.

Figure 6.20*: C10 'Cyclists in Westminster' (Westminster, Soho and Edgware Road)*

D11 – 'Dual carriageway joy riders'

This cluster has 33 records (7.7% of the total), with an average distance of 0.9, signifying uniform characteristics of this cluster. It is most similar to cluster D2. The volume per record is 2.4 which is small and indicated non route hotspots. The cluster demonstrates an over representation of collisions which have three casualties and that occur in the early hours of the morning between midnight and 7 am with a higher proportion occurring nearer to 7am. The low index score of pedestrians and cyclists indicates a strong 'vehicle only' pattern to the cluster. Membership of this cluster suggests a propensity for careless driving on main roads and dual carriageways, particularly on Friday evening and early Sunday morning (Table 6.21). The count data suggest that the hotspots occur on main roads because of the low number of pedestrian crossings, bus stops and cycle lanes.

HIGH			LOW		
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Casualties = 6	316	4.72	Unknown	47	14.57
Other weather	223	12.12	Schools	47	4.37
Time 0400 – 0659	191	25.52	Cyclists	45	75.69
Casualties = 5	191	8.24	Number	8	83.02
Time 0000 – 0359	185	33.83	Number cells	8	82.83
Casualties = 3	171	71.81	Tube stations	0	4.03
Number of speed	156	7.02	Snowing	0	3.5

Table 6.21: *Characteristics of Cluster Type D11*

The hotspots in this cluster are generally very spread out across London. There are equal numbers of hotspots occurring in Greenwich (4), Haringey (4) and Croydon (4). There is a low propensity of hotspots at road junctions and there generally appears to be a large number of hotspots occurring in Zone 2 towards Outer London. Figure 6.21 shows a pattern of hotspots in Outer London near Tottenham, Edmonton and Hornsey, and reveals the pattern of large A roads and dual carriageways. Figure 6.21, a map of Brixton and Camberwell also portrays the increase in hotspots from cluster D1 on A roads.

Figure 6.21: *D11 'Dual carriage joy riders' (Camberwell)*

D12 – 'Main road multiple victim collisions in Outer London'

The final four clusters are similar insofar as they all have a high level of variance which is unexplained. Cluster D2 has the highest number of records, 54 which is 12.65% of the total. The volume per record is very small at 2.55, indicating an average hotspot size of three cells. The spanning tree reveals that this cluster is most similar to clusters 6 and 11. There is a high index and variance score for two casualties being involved in a collision. The results also show an index score of 127 for serious collisions (variance of 27.45), indicating a higher than average proportion of collisions resulting in a serious injury. This cluster has a higher proportion of collisions occurring in the dark (index score = 119, variance score = 55.66). Therefore the nature of this cluster would be collisions involving possibly a high number of vehicles resulting in a serious injury of two or more casualties during the hours of darkness (Table 6.22). As with D1 this cluster demonstrates an under representation of cyclists and pedestrians. The differences are that this cluster exhibits no pattern with respect to day of the week. The hotspots have collisions resulting in serious injury and there are more vehicles involved. From the count data this cluster consists of 18% of all hotspots with speed cameras indicating a high proportion of main road sites. The count data also show that cluster D2 constitutes 20% of the total collisions with five vehicles, a significant proportion.

HIGH			LOW		
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Vehicles = 5	433	12.57	Schools	58	4.37
Snowing	382	3.5	Tube stations	49	4.03
Casualties = 5	333	8.24	Vehicles = 1	48	57.42
Number of Speed	305	7.02	Pedestrians	47	62.11
Casualties = 4	213	6.95	Cyclists	45	75.69
Vehicles = 4	199	5.2	Number	7	83.02
Casualties = 2	191	45.78	Number cells	7	82.83

Table 6.22: *Characteristics of Cluster Type D12*

Type D12 has most of its hotspots in Outer London. The boroughs with the highest numbers of hotspots are Hounslow and Barnet (7) and Enfield (6). From Figure 6.22, the map of Hendon shows a high propensity for hotspots located on dual carriageways. Similarly, the hotspots near Enfield show an increased number occurring along dual carriageways, with low propensities occurring at junctions.

Figure 6.22: *D12 'Main road multiple victim collisions in Outer London' (Hendon)*

D13 – 'Sunday afternoon multiple casualties'

This cluster consists of 34 hotspots which is 7.96% of the total number of records. The average distance is 1.2, which as with most of the other clusters makes it uniform. However, the error of this cluster is higher than average, and it is the second highest out of all the fifteen clusters, with a percentage of 6.96%. The volume per record is 3.9, indicating an average of four cells per hotspot. D3 is most similar to D14 and E15. The characteristics of type D3 consist of vehicle only hotspots on major roads. Because of the high index score of cycle lane length it is feasible that the majority of these hotspots occur on A-roads and dual carriageways rather than smaller roads. The count data reveal that this cluster has 21% of all collisions which account for a total of four injured casualties. This is significantly above the average and highlights the high proportion of casualties in this hotspot that result in slight or fatal injuries (Table 6.23).

HIGH			LOW		
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Casualties = 4	453	6.95	Pedestrians	47	62.11
Casualties = 6	336	4.72	Snowing	37	3.5
Fine with wind	224	6.21	Cyclists	27	75.69
Vehicles = 4	211	5.2	Tube stations	26	4.03

Raining with wind	186	7.77	Schools	15	4.37
Length cycle	173	10.29	Number	13	83.02
Casualties = 3	172	71.81	Number cells	13	82.83

Table 6.23: *Characteristics of Cluster Type D13*

The locations of the hotspots belonging to D13 are spread out across Outer London with hotspots occurring in Barnet (3), Brent (3), Enfield (3), Havering (3), Newham (3) and Waltham Forest (3). The map of Willesden (Figure 6.23) shows that there is propensity for dual carriageway sites. This pattern is also more likely to occur at road junctions.

Figure 6.23: *D13 'Sunday afternoon multiple casualties' (Willesden and Wembley)*

D14 – 'Risk taking early risers'

Despite this cluster having the highest error compared to the other 14 clusters, it consists of 35 records (8.20% of the total) and has an error of 7.54%. The average distance is 1.4, making the hotspots consistent in character. Finally the volume per record is 3.5 which is quite small. This cluster exhibits similar patterns with cluster 13, insofar as the table above lists the variables by highest scoring index scores. However the variance of these attributes is low. Therefore it has been necessary to look further at the data and at variables which although not having the highest index scores are nevertheless still above average and have a high variance. Both cycle length and road length have a particularly high index score (178 and 133 respectively). There are a high proportion of hotspots where collisions are occurring between 4am and 7am. There is a strong over representation of collisions within these hotspots involving 3 casualties which has a particularly high variance score of 71.81 (Table 6.24). These hotspots predominantly occur on weekdays, early morning, with a high number of vehicles (in vehicle only collisions). The count data reveal that this cluster consists of 30% of all collisions involving five vehicles and 24% of those which involve four vehicles. Both of these results support the index scores. Nearly 15% of all collisions involving five casualties are accounted for in this cluster, it is a much smaller and insignificant percentage for six casualties.

HIGH			**LOW**		
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Vehicles = 5	641	12.57	Tube Stations	55	4.03
Vehicles = 4	527	5.2	Vehicle = 1	44	57.42
Snowing	321	3.5	Cyclists	37	75.42
Raining with	309	7.77	Pedestrians	27	62.11
Vehicles = 3	258	23.88	Number	10	83.02
Casualties = 5	243	8.24	Number cells	9	82.83
Casualties = 6	199	6.95	School	0	4.37

Table 6.24: *Characteristics of Cluster Type D14*

Hotspots in this cluster are more likely to occur between inner and outer London in boroughs such as Tower Hamlets (5), Ealing (4), Greenwich (4) and Hillingdon (4). Generally these hotspots are more likely to occur at large junctions. For example, Figure 6.24 of Poplar in East London shows a number of hotspots, at large junctions. Figure 6.24, shows a similar pattern further east towards East Ham, except that these hotspots occur on the dual carriageway. In summary therefore, this cluster is characterised by early morning commuter 'pile ups' involving a significant number of vehicles, at large junctions on the periphery of Inner London.

Figure 6.24: *D14 'Risk taking early risers'*

E15 – 'Sunday morning pedestrian risk takers'

This cluster is characterised by its small size, with just two hotspots and a very high average distance of 35.5 indicating a heterogeneous cluster. This suggests two very different hotspots. The average number of grid cells per hotspot is 2 indicating small isolated hotspots. This cluster has unusually high index scores, as seen for the first variable of collisions involving three casualties, which has a corresponding high variance score. This cluster has a significant number of hotspots with a high number of casualties, given the index and variance scores of both casualties per collision, the number of vehicles involved (2) and the number of casualties involved (3). Other variables with high index scores include fine with no winds and pedestrians each with being more than twice as likely to occur in the hotspots in this cluster (Table 6.25). These variables also have high variance scores, indicating a confidence in their involvement in the possible cause of the hotspot. The results also reveal the over representation of collisions in this cluster occurring on a Sunday (Index score = 163, Variance = 40.7).

HIGH			LOW		
Variable	**Index**	**Variance**	**Variable**	**Index**	**Variance**
Casualties = 3	4423	71.81	Vehicles = 5	0	12.57

Vehicles = 4	528	5.2	Snowing	0	3.5
Casualties per	512	63.63	Rain with wind	0	7.77
Schools	492	4.37	Casualties = 5	0	8.24
Casualties = 4	462	6.95	Fine with wind	0	6.21
Vehicles = 2	378	49.23	Severity = fatal	0	9.16
Number of speed	359	7.02	Casualties = 6	0	4.72

Table 6.25: *Characteristics of Cluster Type E15*

These two hotspots occur in Redbridge and Islington, which indicates that this cluster has no inner or outer London pattern. In summary this cluster demonstrates characteristics associated with Sunday morning pedestrian collisions with high number of casualties being injured in each collision (Figure 6.25).

Figure 6.25: *E15 'Sunday morning pedestrian risk takers' (Shoreditch)*

6.7 Campaigns

From the analysis of the clusters, it is apparent that there are over representations of vulnerable road users in a majority of the clusters. These vulnerable road users include pedestrians, cyclists and children. However, the comprehensive results from the clustering indicate that although these vulnerable road users are at risk, the patterns of risk vary according to spatial location and temporal changes. The management of road safety across London lies with a number of different agencies and partners who maintain a universal goal of reducing casualties and educating the vulnerable on their risk. In November 2001, the Major of London outlined 'London's Road Safety Plan' which was based on the national road safety document 'Tomorrow's roads: safer for everyone' (Department for Transport 2000). The London plan in conjunction with Street Management of TfL outlined reduction targets for London over the next ten years.

There are two main techniques by which road safety education and campaigns are instigated. The first way involves the London Accident Analysis Unit (LAAU) identifying sites for further treatment and then sends this information to the specific boroughs across London to manage the sites in their own way. Secondly, there is an organisation called the Pan London Road Safety Forum, which is a partnership – led scheme involving a variety of stakeholders including the police and other emergency services. This Forum has two sub groups; 'Campaigns and Education' and 'Research and Development'. Its aim is to develop London wide schemes which can help the prevention and reduction of collisions.

The majority of campaigns across the boroughs involve targeting pedestrians, cyclists, children, teenagers and powered two wheeled vehicles in terms of road safety. These road safety interventions include television advertisements, posters, presentations and leaflets. However, as the results have shown the frequency of vulnerable road users varies considerably from borough to borough and across time and space. Whilst this strategic approach addresses the overall high risk casualties, there is a general neglect for a strategic approach to the temporal and spatial disparities that occur. In comparison the Metropolitan Traffic Operational Command Unit manages the tactical day to day traffic collision investigation and crimes on the road such as drink driving, speeding, seat belt wearing and use of a mobile phone whilst driving.

In terms of tactical and strategic campaigns and education the cluster results offer an important insight into the way campaigns and education could be structured across London. For example Type A2 suggests a specific spatial pattern linked to high streets (notably Catford, Lewisham, Streatham, Wood Green and Tottenham) in specific areas occurring between midday and three o'clock signifying a daylight shopping/town centre phenomenon. This pattern would utilise from pedestrians of all ages interacting with the road environment whilst shopping. By concentrating on the mobility of people and the difference journeys people make every day could play an important strategic role in people's awareness of their collision risk. It is important to remember, however that this clustering model concentrates on finding patterns of road collision hotspots relating to circumstantial and environmental risk rather than behavioural risk. It is impossible to know in this instant the state of mind of the driver or casualties and how that can be minimised.

Type A3, which has a small number of hotspots concentrated near to leisure facilities, involving pedestrians suggests that monitoring of other leisure centres should be undertaken as well as pedestrian engineering solutions such as gaited rails at the side of the roads. In the same way Type A5 has a strong occurrence of hotspots in the Camden area involving pedestrians late at night. This is a prominent pattern revealed in the results, whereby there are a large number of hotspots occurring late at night, or early in the morning, involving pedestrians who have possibly been using the road environment in an unsafe manner. In terms of education and campaigns there are none which address this high risk, predominantly Central London, population.

In comparison, type A8 involves a type of vulnerable road user who is being exclusively targeted by campaigns. In general, the London road safety campaigns focus on child cyclists, which from the results from this clustering can be seen to be an important type of vulnerable road user. However, what these results also show is the high proportion of commuting cyclists who are involved in collisions. Within a country that has not made wearing cycle helmets an enforced law, there is a clear requirement to target the overall cycling population, particularly within Central London. This is of specific importance with regard to type C10 which demonstrates an overall trend of cyclists involved in collisions in a very busy road environment, including hotspots such as Oxford Street, Tottenham Court Road and Edgware Road. With the vision by the current Mayor of London, Ken Livingstone to make London a 'world class cycling city' (www.lcc.org.uk 2005), a stronger road safety campaign needs to be forthcoming in order to reduce the number of casualties.

Reducing the number of casualties on London's roads is a vast and challenging mission. However, the more information and understanding there is about the types of collisions that are occurring, and particularly where they are occurring means that better campaigns and education can be rolled out. The results of this clustering highlight this information and potential for its use in these reduction schemes.

6.8 Conclusion

The intention of this chapter has been to profile road collision hotspots in London. The aim of profiling the hotspots was to identify patterns across time and space, in a way which would assist the reduction and prevention of high density collision areas. The classification of road collision hotspots in road safety still remains an important and under developed theme. This chapter has attempted to highlight how the clustering process works and the interpretation of the cluster types and groups. The cluster types have led to some interesting patterns across time and space, particularly the strong divide of clusters between those that involve pedestrians and cyclists and those that do not. Predominantly the cluster types involving pedestrians and cyclists occur in Central London whilst vehicle only cluster types are more likely to occur on the larger more arterial roads around Central London. By using the STATS19 dataset variables and environmental indicators it was possible to develop a more robust description of the clusters, the most probable times of day, the day of week, and the types of scenario which are likely to occur in the cluster to make the different hotspots evolve. The spatial analysis of the cluster aims to portray the likely type of areas in which the hotspots occur. These locations include particular boroughs which experience a high number of a certain type of cluster or certain roads which have hotspots from a specific cluster which run along that particular network. This type of information leads road safety professionals to a better understanding, not only of the types of hotspots but their patterns across London, whereby, for example boroughs such as Barnet and Hillingdon may understand commonalties in their similar patterns of cluster types. Important policy implications will also be enforced when the UK police forces merge from 43 forces down to 12. The consequences of this and the introduction of neighbourhood policing in 2006 will have implications for the management of road safety and data management.

As with any classifying process there is a limit of the number of attributes that may be included in the process, and limitations of the chosen attributes. The limitations of the variables that are input should not be taken to undermine the importance of the results, but merely cite the areas that would be modified to enhance the usability of the clustering outcome. The ongoing increase in the numbers of collisions involving motorcycles suggests that desegregation by vehicle type would have increased the knowledge of powered two wheeled vehicles in the cluster types. In conjunction with this, the numbers with each age group involved in the collisions either as drivers or victims would have facilitated a more accurate understanding of the cross section of society involved. When analysing the age group in each cluster from the count data it was clear that the most at risk age group in every cluster were aged between 25 and 40.

The clusters created in the clustering process fulfil an important role in understanding the stronger and weaker attributes involved in the hotspots. As with any classifying process, the aim is to understand the stronger collision types occurring in the cluster. The clustering process takes the majority explanation of the hotspots; however, not all the hotspots in each cluster are likely to fit the description.

One of the main reasons for using a clustering methodology to profile the high density hotspots in London arose because it is challenging and largely ineffective to model individual collision circumstances. In a similar way, geodemographics is used because it is impossible to model individual purchasing decisions (Openshaw and Alvanides 1999). In order to be cost effective and parsimonious, there needs to be a high number of collisions in any one area that can be managed. The analytical processes described here make it possible to identify collision likelihood, based on previous events, and without any bench marking studies. As such, this chapter has identified the main cluster types that are to be found in the London area.

CHAPTER 7

A ROAD COLLISION TYPOLOGY

7.1 Introduction

The relationship between where people live and where people are more likely to experience a collision has never been formally investigated in so much as there been investigation into how far from home people are when they have collisions but the originality of this thesis lies in the analysis of people's lifestyles and geodemographic characteristics. Many studies have researched the distance between home to collision (see Scottish Executive 1999), and they have concluded that the majority of collisions occur close to the place of residence (between 0-2km). In the previous chapters there has been an attempt to define and investigate the characteristics of these locations, that is, the collision location and the place of residence of the driver or casualty. The World Health Organisation's global road safety report (WHO 2004) recognised that the socio economic lifestyles of people can pose as a risk factor for being involved in a collision. This poses a public health problem because every road collision results in the use of public health facilities. Previously road collisions were seen as 'random acts' uncontrollable by preventative methods. However in recent years there has been a shift to viewing road collisions as a 'human – made' problem. This chapter aims to merge the previous spatial location descriptions to link together the hotspot typology and the geodemographic propensity patterns for London in order to build a typology to determine which geodemographic types are more likely to be involved in certain collision hotspots within London. Road collisions are often seen as a health risk (road safety was the theme of the World Health Organisation's 'World Health Day' in 2004) indicating a shifting emphasis of research towards viewing road collision risk as a preventable if not manageable risk within society. It is the view of the WHO that the risks of being involved in road collisions are manageable because these risks can be distinguished within society and therefore managed. Risk factors would include socio-economic status, travel patterns, personality and behavioural characteristics of the road user and of the road network itself. Therefore the importance of understanding the characteristics of the people likely to be involved in high density collision areas should be a fundamental aim for road safety. Because the UK police are not, by law, required to collect postcode information for drivers and casualties, the research and its role in facilitating road safety plans has been limited. While the WHO report identified that the poor and vulnerable members of society are more likely to be casualties of road collisions (WHO, 1999), this does little to pinpoint the exact nature of the exposure to road collision risk for individual people with regards to spatial and temporal patterns of exposure. This chapter aims to identify the propensity of risk for each geodemographic type based on the nature of collisions that they are most likely to be involved in. This combines the work from chapters four, five and six by linking the two locations together as well as the level of risk experienced by the people involved in collisions in the hotspots.

7.2 The coincidence of 'hotspot' and 'place of residence'

A person's postcode is key to understanding the environment in which he or she lives. Targeting people who are at risk from certain 'types' of collision should be an important crux of road collision management and reduction. In the previous three chapters there has been an attempt to define and profile the two locations from environmental, temporal and spatial standpoints (the 'hotspot') and a socio economic view (the 'home' location). There have been various studies to understand the influence of 'home' background upon certain people's collisions such as young adults (Murray 1998), however there has been little or no research into the link between the socio economics of an individual's postcode and the type of collision he or she is most likely to be involved in. Whereas in recent years the emphasis has been upon child pedestrian road collisions and their socio economic backgrounds, this chapter focuses not only on casualties under 15 but also casualties and drivers from all different age groups. This means that it is possible to distinguish the possible patterns of age propensity which may or may not occur within the hotspot types.

7.3 Classification technique

The first stage of the analysis was to extract both the casualty and driver details for each collision Group and Cluster. This was achieved using ArcGIS to analyse the Cluster and Group polygons created in Chapter Six. It was possible to select the collision details of the drivers and casualties which occurred in each polygon. Because there was usually more than one polygon for each Group or Cluster, the driver and casualty information was aggregated to form two datasets for each Cluster and Group (driver and casualty). Using the count data for the number of people in each Mosaic Type for each collision Group and Cluster, it was possible to calculate the index scores and therefore to understand the road collision risk propensities of each Mosaic Type in the collision Groups and Clusters. Then, using the age count scores for both the collision Groups and the Clusters it was also possible to work out the index scores of each age group within each collision Group and Cluster. Therefore the outcome will present two methods for determining road user risk. One would be able to look up the Mosaic Type a person belongs to and establish the risk propensity for being in a certain type of collision at a particular location. Or one would identify the spatial collision hotspot first and then determine its high risk casualties and drivers and which Mosaic Type they would be more likely to come from. It is important to note that the Collision Groups and Clusters (collated from the previous chapter) are based on collision location only and the fundamental aim of this Chapter is to determine the likely residential patterns of the people involved in these Collision Groups and Clusters. The next section gives detailed descriptions of the pattern of propensities found in each Group and Cluster with regards to Mosaic Type.

7.3.1 How the index scores were calculated

$$\frac{\text{Total casualty count (for each cluster and group)}}{\text{Total individual population for London}} \times \text{Individual Mosaic count (for each Mosaic type)}$$

$$\frac{\text{Total casualty/driver count for particular age group (group and cluster)}}{\text{Total age group populations for London}} \times \text{Individual Mosaic count for particular age group within cluster/group (for each Mosaic Type)}$$

Equation 7.1: *Index score equations for road collision typology and age indexes*

The first equation in Equation 7.1 shows the index score calculation for the overall index scores for the Mosaic Types. However this gives an overall indication of the risk that is experienced. The second equation shows the calculation for determining the index scores for the specific age groups within each cluster to establish the risk to certain age groups.

Collision risk and index scores: what do they mean?

Collision exposure can be interpreted very broadly and a wide variety of exposure measures have been used in the field of road safety, ultimately the choice of exposure measure is based upon the elements within the study that are under investigation. For the purpose of this study exposure has been interpreted in it own unique way. The index scores are a measure of exposure to road collisions based on a person's socio-economic and lifestyle characteristics. However it is important to think about the idea of exposure and risk as it plays an important theoretical role in the study of road safety and is ultimately a very ambiguous concept.

Chipman *et al* (1995) suggested that risk exposure can be measured in terms of population, driver distance or driver time. Chipman *et al* (1995) argued that these attributes are most appropriate measure of risk because travel and therefore exposure occurs within all these three dimensions. The following table lists examples for each class of measure. Throughout the literature the most common exposure measure is distance travelled and time to determine exposure to risk. However this was not the most appropriate for this study. Firstly, distance and time travelled can only be used to determine driver risk exposure rather than all other road users. Secondly it is largely impossible to determine distance and time travelled for a large scale analysis of road safety. This attribute can only be attempted to be collected and used on very small scale studies whereby some qualitative analysis is conducted with all drivers within the study area. This method was not possible for this thesis (see Figure 7.1).

Exposure Dimension	Exposure Measure
Space	• Driver/vehicle kilometres • Passenger kilometres
Time	• Driver/vehicle hours • Passenger hours • Average Daily Traffic (ADT)
Population	• Number of licensed drivers • Number of registered vehicles • Number of trips • Number of passengers • Population

Figure 7.1: *Example of different exposure dimensions and how they might be measured* (http://people.hofstra.edu/geotrans/eng/ch7en/meth7en/ch7m4en.html June 2006)

However in road safety analysis exposure is considered to be key information. Surely when in road safety we talk about risk what we really mean is exposure? The literature inter tangles these concepts together: however, it is possible to deem that actually they are the same measure. Exposure is just the type of risk experienced by road users. People have higher levels of exposure than others and this can be measured very differently. Using socio economics and lifestyle as a measurement of exposure is not the only way or the best but it is a measure which previously has been unexplored and misunderstood and can be applied at a large scale if the appropriate data are known (postcode information). Therefore depending on the scope of the analysis, some exposure measures maybe more or less relevant. Second, some measures of exposure are simply not available.

Therefore with all this ambiguity of exposure data, what it measures and how it is collected there are inevitably questions about the reliability and usefulness of the data. It is easy to conclude that road danger cannot easily be quantified, and that the most we can expect from any scientific road collision study is a 'best guess' scenario. Adams (1995) summarises that with regards to road collisions it is not clear how exposure might be measured. He asks the question; how would one measure the duration of exposure, its intensity and quality (Adams 1995).

What Adams (1995) goes on to think about is the changing risk and therefore exposure and how it differs with regard to cultural influences. This concept of cultural filtering is of particular importance for this study, as the Mosaic Types give indications of the different cultural societal trends which are present (for example 'South Asian Industry'): even if the cultural link is not explicit in the name of the Mosaic Type, the label suggests cultural inclination. Adams (1995) discusses the idea of a 'risk thermostat' by which everyone perceives their own risk, and how this varies between groups of people, cultures and obviously over time. These cultural filters select and construe evidence to support established biases. There are some threats on which all cultures can agree for example, all drivers slow down when they come to a sharp bend in the road, and they are in general agreement about the nature of the risk. However some subcultures would slow down more than others. Adams (1995) argues that this filtering process works both directly (through our five senses) and indirectly (through reactions to stories, news reports, statistics and research). These are in effect pre filters, by which no government or research institution can gather even a fraction of this evidence. Adams (1995) summarises the road safety policy of the western world (including the UK) as having just two aims, of making motoring safer for motorists and getting everyone else safely out of the way: his view, therefore, is that there is little need to research into the activities of those who are displaced.

What is also important and not studied in greater depth is the interpretation of different cultures to the risks on the road they face. Can people understand the government's statistics of 'risk to car occupants per vehicle mile travelled'? Different people will respond to road safety measures in different ways and this ultimately has not been explored. This research is only one cog in the machine of trying to investigate this phenomenon further. By understanding that different people have different levels of exposure to risk we can start to think about how this might be managed. Of course, some exposure can be obviously related to Mosaic Types. For example, Type H46 'White Van Culture' indicates a lifestyle which incorporates a larger proportion of time spent in the road environment that other Mosaic Types. Therefore an over represented level of risk exposure can be expected, but what we are interested in here is the level of exposure, compared to other Mosaic Types and most importantly, where this risk is encountered. No one knows how risk varies according to sub groups, because detailed data on activity patterns are not collected for extensive and representative samples of people. This study presents itself as a valuable exploratory investigation into the 'what is' analysis of variation in risk. Knowing they are likely to be over represented is one thing but understanding the spatial pattern is also important. The next few sections concentrate on summarising the Mosaic Types which are associated with each collision group and cluster, the final section focuses on trying to disseminate what can be learnt from this in a wider context and how it can be interpreted.

7.4 Collision Group analysis by Mosaic Type

The risk associated with different age groups and their likelihoods of becoming involved in a collision has been well documented over recent years (Massie *et al* 1995 and Ryan *et al* 1998, Abdel-Aty *et al* 1998). The previous discussion in the literature review (Chapter Two) outlines in more depth the findings of the most recent and relevant research relating to this thesis. This chapter, however, addresses the likelihood of different age groups within each collision Cluster and Group becoming involved. Each Cluster and Group was disaggregated by age group (under 15, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64 and 65+) and the counts of each age group were used to disseminate index scores using the overall total of London's population in each Mosaic Type as the base population for the index scoring. Age was not included in the Clustering algorithm because of the possible and potential complications of driver and casualty and the increased level of data may have compromised the accuracy of the clustering process. However this is not known precisely. It also would have added nine more variables. Age is extremely important in gauging risk; road safety campaigns focus on children for example and car insurance premiums usually focus upon age as a direct indicator and as a correlate of driving experience. Therefore this next section discusses the patterns of risk for each age group and associated Mosaic Types.

Collision Group A 'Central London Pedestrians' is characterised by central London pedestrian collisions. The results are determined by ranking the Mosaic Types by percentage, first by the percentage of casualties (overall in London) in each Mosaic Type and then by percentage of the total London population which belongs to each Mosaic Type. In Mosaic Group A casualties Type D27 'Settled Minorities', disaggregated by age, shows that all the age groups apart from those under 25 are more likely to be involved in Collision Group A. Analysing by age group, the under 15 age group is most over represented in the Type A01 'Global Connections', and Type A02 'Cultural Leadership' indicating that these children could be involved in a disproportionate number of collisions. In the age 15-19 casualty category for this collision Group, both D26 'South

Asian Industry' and F36 'Metro Multiculture' have over represented risk exposure. This indicates a tendency for teenagers of ethnic origin to be more likely to be a casualty in this group. In the 20-34 age groups, Type E29 'City Adventurers' are over represented amongst those that are casualties in collisions, as well as F36 'Metro Multiculture' and E28 'Counter Cultural Mix'. In Collision Group A, the people who are likely to be involved under the age of 44 are over represented in a range of Mosaic Types, particularly those from ethnic minority population types. It is almost inevitable that there will be some under representation in some Mosaic Types and over representation in others, over all. However the older people likely to be casualties are more polarised and come from the Mosaic Types with the highest percentages of the total London population. Drivers in this Group form different population proportions, for example just over 12% of the driver population is made up of H46 'White Van Culture': however, in Group A they are not over represented in any of the age groups. The patterns remain similar for drivers and casualties in this Group. Although we expect H46 'White Van Culture' to have a high index score and therefore higher risk exposure, this indicates that the roads on which H46 'White Van Culture' are more exposed are Central London roads as well as Outer London roads where this Mosaic Type is more likely to live.

Collision Group B 'High density vehicle damage' where the majority of the collisions that occur have a very high spatial density and occur at the weekend involves a very different pattern of age groups and Mosaic Types. The under 15 age group is also over represented in Type A01 'Global Connections' with an index score of 115. In the general Mosaic Type index scores, Mosaic Type C15 'Close to Retirement' revealed an index score of 127. This Collision Group however contains only one Cluster and is therefore quite polarised in its pattern of age propensities and many of the index scores are very high. Therefore this means that no concise and accurate conclusions can be drawn from this group of drivers.

Collision Group C, 'Cyclists in danger', comprises only one Cluster; however this Cluster is characterised by its unique position. It covers a large spatial proportion of central London, including Oxford Street, Edgware Road and Tottenham Court Road with a strong propensity for cyclists to be involved. In the 15-19 age group, there a strong propensity for people from A03 'Corporate Chieftains' to be involved in a collision: the pattern for this Mosaic Type is also applicable to the 20-24 year old age group. Type A01 'Global Connections' has a strong propensity for people aged 45 and over in this Group: this is to be expected given the location of where they live, but not in terms of road collision exposure based on the argument that they driver more safe upmarket cars and therefore less likely to be involved in a road collision. This finding therefore introduces a different pattern seen in the general patterns of risk. The casualties in this Group have a tendency to involve people from both affluent Mosaic Types, and Types which have greater financial constraints; however the majority of the people would live and possibly work in Central London, rather than originating from the suburbs of London. Therefore the casualties and drivers in this Group are likely to be from diverse backgrounds making a definitive pattern difficult to distinguish. The drivers in this group follows the pattern of the casualties, indicating therefore that the people involved in collisions in this group are just as likely to be involved as a casualty or as a driver. It is also important to note that the majority of casualties in this Group are cyclists who are classified as drivers. Overall Mosaic Type index scores for casualties indicate high exposure for Types A01 'Global Connections', 'F36 'Metro Multiculture', and E28 'Counter Cultural Mix' which are generally very different Mosaic Types, with different cultural and socio economic characteristics.

Collision Group D, 'Multiple main road collisions', has a high proportion of Mosaic Types with over represented involvement in collisions including H46 'White Van Culture' and C18 'Suburban Sprawl'. From the risk propensities indicated by the index scores for this group, it is clear that the people most likely to be involved in collisions from Group D are residents of the more suburban areas of London. For example Type C18 'Sprawling Subtopia' shows above average index scores for Collision Group D as well as a considerable increase for H46 'White Van Culture'. Type C20 'Asian Enterprise' also has a higher than average propensity to be involved in a collision. The collisions in this Group are more likely to be on main roads than in the suburbs and to have high numbers of casualties involved. 0-19 year olds in this group are more likely to be 'H46 White Van Culture' and B12 'Middle Rung Families'. Drivers in this Group have a preponderance to be from Type C20 'Asian Enterprise', with the over representation of this Mosaic Type peaking for the age group 15-19. There is a tendency for a Mosaic Type to be over represented in all age groups, for example, C20 'Asian Enterprise' is over represented in every age group. This pattern also applies to H46 'White Van Culture' and D27 'Settled Minorities'. However there are a number of Mosaic Types which have very different risk propensities across the age groups. These include for example A03 'Corporate Chieftains': this Type is over represented in age group 15-19 (possibly the children of the adults in the Mosaic Type), and there is also an over representation for people aged 20-24 and 45-64 for drivers.

Finally Collision Group E, 'Weekend risk takers', comprises only one Cluster and has very high index scores, however this is the result of the small numbers of collisions and over exaggeration of the results. The small number of clusters and drivers and casualties makes it difficult to draw any firm conclusions. Overall, there are no over represented index scores for drivers however there are for casualties.

Overall the group index scores for each age group have indicated that the patterns on a wider scale vary between Inner London and Outer London. The first two groups are predominately associated with Central London, with Group C evoking an original and unique insight into the collisions which occur in Central London. Finally the last two Groups have very different patterns and are associated with more suburban London patterns of involvement, characterised by the type of roads and the Mosaic Types more likely to involve in these two Collision Groups. The Clusters portray a different and more detailed outline of the patterns of age groups.

Table 7.1 shows the Group index scores for all the Mosaic Types (casualties) which make 1% or more of the total London population (Mosaic Types making up less than 1% have been omitted from this table but were included in the analysis).

Mosaic Type	London	London %	Casualty %	A	B	C	D	E
D27 Settled Minorities	843155	11.38	12.7	121	97	78	128	118
F36 Metro Multiculture	912740	12.32	12.3	171	102	161	105	27
E28 Counter Cultural Mix	675251	9.12	8.2	158	230	189	65	37
C20 Asian Enterprise	478737	6.46	6.7	86	47	59	146	174
H46 White Van Culture	348466	4.70	6.2	71	49	42	182	119
C19 Original Suburbs	467482	6.31	6	65	42	51	97	71
E30 New Urban Colonists	492125	6.64	5.1	107	172	98	66	0
C18 Sprawling Subtopia	273904	3.70	4.1	58	83	43	122	365
E29 City Adventurers	432742	5.84	3.9	117	144	164	51	38
D21 Respectable Rows	188055	2.54	3	68	15	52	100	487
A02 Cultural Leadership	344224	4.65	2.9	57	90	79	69	0
A01 Global Connections	393595	5.31	2.2	56	115	154	34	63
A03 Corporate Chieftains	164708	2.22	1.9	41	34	93	63	51
C15 Close to Retirement	88796	1.20	1.7	54	127	51	134	94
B12 Middle Rung Families	85951	1.16	1.7	68	0	53	176	290
J52 Childfree Serenity	147971	2.00	1.6	62	57	54	59	112
E32 Dinky Developments	92444	1.25	1.6	75	61	89	134	270
A05 Provincial Privilege	141025	1.90	1.5	43	40	29	74	0
D26 South Asian Industry	85644	1.16	1.2	139	66	75	121	0
H47 New Town Materialism	48872	0.66	1.1	69	0	75	142	0

Table 7.1: *Index scores for Casualty Groups*

Table 7.1 shows the index scores for each Collision Group, indicating strong differences in composition with regards to Mosaic Types. Groups D and E both have high index scores for Type C20 'Asian Enterprise': however, Group E only consists of only one Cluster, and this has a small number of casualties and drivers, indicating that the results for this group are likely to be largely over exaggerated. Group D however, has a high index score for 'White Van Culture', and Group D is characterised by experience of multiple collisions early in the morning. Groups C and D also have high proportions of Mosaic Type E29 'City Adventurers' which are more likely to be involved in a collision in both those types of collisions found in Groups B and C.

Mosaic Type	London	London %	Driver %	A	B	C	D	E
D27 Settled Minorities	843155	11.38	12.3	123	31	476	125	9
F36 Metro Multiculture	912740	12.32	11.4	155	32	249	97	40
E28 Counter Cultural Mix	675251	9.12	7.5	145	10	157	74	22
C20 Asian Enterprise	478737	6.46	7	91	36	357	157	3
C19 Original Suburbs	467482	6.31	6.3	70	39	402	101	8
H46 White Van Culture	348466	4.70	6.1	80	25	363	148	4
E30 New Urban Colonists	492125	6.64	4.9	104	10	219	65	0
C18 Sprawling Subtopia	273904	3.70	4.3	68	12	276	123	1
E29 City Adventurers	432742	5.84	3.7	115	11	115	61	14
D21 Respectable Rows	188055	2.54	3.1	74	44	157	111	0
A02 Cultural Leadership	344224	4.65	3	64	13	191	71	0
A03 Corporate Chieftains	164708	2.22	2.3	52	17	78	96	4
A01 Global Connections	393595	5.31	2.1	59	12	111	40	7
C15 Close to Retirement	88796	1.20	1.8	69	2	75	109	1
B12 Middle Rung Families	85951	1.16	1.8	81	0	71	175	0
J52 Childfree Serenity	147971	2.00	1.6	61	9	120	66	2
A05 Provincial Privilege	141025	1.90	1.6	44	12	212	68	0
E32 Dinky Developments	92444	1.25	1.6	95	5	45	147	0
D26 South Asian Industry	85644	1.16	1.2	139	5	50	143	0
H47 New Town Materialism	48872	0.66	1.1	85	0	29	134	0
A06 High Technologists	42654	0.58	1	72	2	39	113	0
B11 Families Making Good	37249	0.50	1	102	0	17	139	0
B13 Burdened Optimists	31072	0.42	1	132	0	15	233	0

Table 7.2: *Index scores for Driver Groups*

Table 7.2 shows the index propensity for drivers. The index scores in Tables 7.1 and 7.2 are broadly similar with some small differences. For example, Group A drivers have similar risk propensities to those of casualties in this Group. This pattern indicates high index values (123, 155 and 145 respectively) for Mosaic Types D27 'Settled Minorities', F36 'Metro Multiculture' and E28 'Counter Cultural Mix'. One of the fundamental differences between the Driver and Casualty Groups results is that Group B drivers do not have index values above 100 for any

Mosaic Type, indicating that there are no drivers in any Mosaic Types which are more likely to be involved in this type of collision (Group B is characterised by high density collision hotspots usually occurring in the dark on Saturday evenings). Group C, however, (characterised by one Cluster, C10 which is located in Central London) has very high over representation for many Mosaic Types. These include D27 'Settled Minorities', which are four times as likely to be involved as a driver in this type of collision. This is followed by C20 'Asian Enterprise' whose members are four times more likely to be involved in a collision in Central London as a driver. Group D has similar propensities.

7.5 Profiling the 'who' in the collision clusters

Having now begun to unpick the patterns within the Collision Groups it is necessary to understand who (in the London population, in geodemographic terms) is more likely to be involved as a driver or casualty in the Clusters outlined here. The motivation for this lies in differentiating between sections of society who will be more likely to be involved in certain collisions. Obvious uses for this information include better strategic policy for targeting specific neighbourhoods with information about the risks that residents face, and the selective implementation of preventative measures. In this context Fontaine *et al* (1997) created a typology of pedestrians in France. Although the methodology that they developed is different, not least because it has no spatial element, the aims and objectives are similar. Fontaine *et al* (1997) surmise that 'Such [an] education program should be adapted to the pedestrian group, and particularly the age group, that they address, insofar as the problems are different in each category'. In this case, a vulnerable group was chosen (the pedestrian) and investigated further in terms of the different types of pedestrians likely to be involved (Fontaine *et al* 1997). This section of the chapter outlines in depth the patterns of risk associated with Mosaic Types in each Cluster to identify who is more likely to be at risk from certain collisions (see Tables 7.3, 7.4 and 7.5)

A1 'Late night football supporters'

This Cluster has only 9 casualties and 16 drivers recorded as it only had one Cluster in Upton Park near a football ground. The majority of the injured were pedestrians, with many drivers involved in the collisions but not injured. The casualty index was high for Type C20 'Asian Enterprise' with an index score of 193. Type F36 'Metro Multiculture' had an index score of 203. This type of cluster is applicable to many other situations in London, generally linked to the theme of crowd management and road safety. This is applicable to many sporting events in the capital as well as concerts and other large events where there are surges in the numbers of pedestrians using the roads.

Four out of the five casualties are female, of which only one was a pedestrian. Three out of the four women were aged 59-65, indicating a possible high risk age range for this particular Cluster, which is interesting because of the expectation that the Group would comprise young males. The breakdown of Mosaic Types by age group reveals a polarised pattern because of the small number of casualties in this Cluster.

Drivers in this cluster are likely to come from the most 'expected' Mosaic Types, with regards to the Mosaic Types with the highest proportion of the population. These include F36 'Metro Multiculture', D27 'Settled Minorities' and D26 'South Asian Industry'. This Cluster and its associated high risk collision victims is slightly polarised because of the small and distinctive nature of this Cluster. Disaggregating the vehicle type for the drivers in this Cluster, 15 out of the sixteen vehicles were cars and the majority of drivers were males. The age analysis of the drivers produces two distinct categories. The first is drivers aged 24 to 37; the second group aged 47-58, with a significant proportion of these being male.

A2 'Inner City pedestrian risk takers'

Type A2 has approximately 2000 casualties and 3500 drivers involved in the collisions within this Cluster. 900 of the casualties were drivers or riders, followed by 652 casualties being passengers and 545 being pedestrians. This pedestrian count, compared to the other Cluster types is very high and supports the findings of the Clustering results that this type is predominantly focused on pedestrians and their involvement in collisions. Type A2 has the highest percentage of casualties under 16 and casualties over 65, indicating a superfluous vulnerable casualties group relative to the other Collision Clusters. This Cluster reveals a high number of Mosaic Types having an over representation in this type of collision. Another Type which has an over representation of being involved in a collision as a casualty in this Cluster is Type H46 'White Van Culture' which has an index score of 122. This not an Inner London expected pattern based on the patterns found in Chapter Four and the concentric rings around London. Other similar proportions include Types E32 'Dinky Developments' and D26 'South Asian Industry with index scores of 132 and 125 respectively. The latter of these, as with H46 'White Van Culture', is also associated with Outer London rather than Inner London.

The general pattern is of the highest frequencies of collisions experienced by people aged between 20 and 40 years old. More young males are injured than females. The total number of female casualties under 16 is 84, much lower than the corresponding figure of 136 for male casualties. This may arise because males use the roads more which could suggest a target age group for road safety initiatives. This shows a high risk group in this type of Cluster, predominantly focused on shopping centres. A high proportion of passenger casualties is aged 15-27, mostly travelling in cars. There is also a pattern of elderly people aged 70-90 being involved as passengers in collisions where the driver is younger. Overall however in Type A2 there are more females injured as a pedestrian than males, and this pattern is the same for passengers. Injured drivers are three times more likely to be male. This Cluster is characterised in terms of Mosaic Type and age group by a large proportion of the younger age groups being more likely to come from D26 'South Asian Industry'. The 15-19 age group is over represented by H46 'White Van Culture' and D26 'South Asian Industry'. The 25-34 range is dominated by E32 'Dinky Developments' and their over representation increases in the subsequent 35-44 age range. In the older age groups the casualties of these groups are more likely to be from J52 'Childfree Serenity'.

Drivers in this Cluster Type have similar index scores to the casualties. The major difference is the lower index score for Type H46 'White Van Culture'. There is also a considerable drop in propensity of D26 'South Asian Industry'. Overall there are higher numbers of male drivers involved in collisions; however this arises because of the higher overall number of male drivers across London. There is a large proportion of motorcycle driver (between 125-500cc) involvement in collisions. One of the key high risk manoeuvres in this Cluster type is stopping or slowing. The drivers in this Cluster who are younger tend to be from a smaller selection of Mosaic Types. The older the casualty the more diverse the over represented the Mosaic Types tend to be; therefore they are more likely to be from more diverse backgrounds. This is especially true for the 55-64 and over 65 age groups.

A3 'Saturday morning leisure time hotspots'

Cluster A3 is small with only two hotspots occurring near leisure centres at weekends which as with cluster A1 this collision location can be applied to similar areas in London with leisure centres. There are four casualties and ten drivers occurring in this Cluster which makes any definitive conclusion unreachable, but a level of interpretation and understanding of the cluster is useful. All the casualties were drivers/riders, with two of these being pedal cyclists. It is difficult to gauge patterns from such a small number of occurrences. However out of the four casualties, three were represented in the top four Mosaic Types for London; F36 'Metro Multiculture', D27 'Settled Minorities' and E30 'New Urban Colonists'. Three out of the four casualties were male and were in their late twenties to early thirties. The Mosaic index scores between drivers and casualties are very similar. The ages of the drivers fall again into two distinct groups, the first being 20-32. The second predominant age group is 50-55 (two of those being female). The type of people therefore likely to be injured in these areas is young adult males either driving cars or bicycles. The drivers are similarly of young adult age and also older male and female drivers. Two points can be drawn from this cluster, firstly whether or not the casualties and drivers were linked to the leisure centre i.e. whether they were taking use of its facilities. Secondly, the idea that potentially leisure centres should be treated like schools in terms of road safety initiatives, because of the increased number of children and young adults that are likely to use the facilities.

A4 'Sunday afternoon car drivers'

This Cluster and its casualty attributes is characterised by having 2215 casualties and 4196 drivers. This Cluster has a high proportion of under age 16 casualties. It also comprises a moderately high number of over 65 casualties. The casualties in this Cluster have a high propensity of being from Type F36 'Metro Multiculture', D27 'Settled Minorities' and E28 'Counter Cultural Mix'

Of the proportion of F36 'Metro Multiculture' casualties, there is a significant group of males aged 29-34 which appear to have a higher representation within the collision statistics for this Cluster than any other age group. The peak for F36 'Metro Multiculture' female casualties has lower index scores than the males but generally occurs when they are 27-31 years old. Proportions of child casualties (under 16) are high, especially for males; however this trend is represented throughout most of the dataset. Generally with regards to casualty class there is a significant proportion of F36 'Metro Multiculture' children involved in collisions as a pedestrian compared to being a passenger in a car. Compared to F36 'Metro Multiculture', D27 'Settled Minorities' has a more diverse age pattern of its casualties. The females tend to peak between 20-30 whilst the males have three peaks at 24, 31 and 36 with smaller ones in between. Type F36 'Metro Multiculture' is less likely to be over represented as a driver in this hotspot. The index score for drivers is 139 compared to 169 for casualties. Type C20 'Asian Enterprise' on the other hand has a higher index score as a driver at 125. One of the highest index scores however of 149 and a population proportion of 11.38% is Type D27 'Settled Minorities': when disaggregated by age, sex and vehicle type this Mosaic Type shows that a high proportion of males aged 20-40 are more likely to be involved in a collision as a driver. Women from this Mosaic Type have a small elevated peak at the age of 25. The casualties in this Cluster for all age groups tend to be more likely to come from the top three or four Mosaic Types (in terms of %), which can be seen in Table 7.3.

Overall, the drivers of this Cluster are more likely to be aged 27-38; this is a significant peak in the over-all age profile. However this peak is largely accounted for by male drivers rather than females. The female collision risk propensity is more spread over the age groups whereas the male propensity is more pronounced. Drivers are likely to be turning right when involved in a collision. With regards to vehicle type, cars are involved in the highest proportion of collisions within this Cluster. In this Cluster the drivers aged from under 15 to 54 are more likely to come from similar Mosaic Types. The older (55+) age groups have very different patterns. For example these age groups are over represented in J52 'Childfree Serenity' and C15 'Close to retirement'. They have a very different involvement dynamic than the younger age groups with regards to the different Mosaic Types represented at this older age.

A5 'Illicit late night Zone 1 pedestrians'

This hotspot entitled 'Illicit late night Zone 1 pedestrians' has high numbers of casualties and drivers. There are 3316 casualties and 6112 drivers. This collision type has 7.4% of its casualties under the age of 16 and 5.6 % over the age of 65. The casualties in this Cluster are polarised by two types F36 'Metro Multiculture' and E28 'Counter Cultural Mix'. In this Cluster F36 'Metro Multiculture', when disaggregated, has a peak age group between 27-32. In the under 16 age category, there are peaks at the age of 12, 14-16 and 58-62. While there are evident increases in casualties aged 21 and then again from 25-30, this is consistent with the Cluster pattern that younger people are using the road environment irresponsibly late at night. With regards to the class of casualty, there are high numbers of those under 16 involved in collisions as a pedestrian; this peak continues to the age of 21, which is in keeping with the pattern of the Cluster, that it would affect young adults indulging in anti social behaviour. The age for driver casualties peaks between 24 and 29 years old. Females from Type F36 'Metro Multiculture' aged 4-16 have a higher proportion of casualties than their male counterparts. Overall, the casualties in this Cluster are more likely to be 26-32 years old, and although those involved are likely to be drivers there is also a high proportion of pedestrians in this age group. There are equal numbers of male and female pedestrians, although a difference between the sexes becomes more distinct when analysing passenger and driver casualties, where it is apparent that there is a preponderance of male casualties. Under 15 casualties in A5 are polarised and likely to come from the top three Mosaic Types (see Table 7.3). The middle age groups (from 20 – 34) are more likely to be from C18 "Sprawling Subtopia' and E29 'City Adventurers' The 55-64 age group has a more widespread over representation in an increased number of Mosaic types, including A03 'Corporate Chieftains'.

Although the drivers in this Cluster generally have similar index scores to the casualties, the one major difference is the markedly higher collision propensity for Type E29 'City Adventures', who have an index score of 128. This Mosaic Type is regularly identified as under represented in collision involvement. This Cluster is marked by the over representation of male drivers involved who are aged in their late twenties and early thirties. Drivers in this Cluster have a more polarised socio economic background indicating that drivers in this Cluster are more likely to be from a small number of Mosaic Types. For example, E29 'City Adventurers' has an over representation in nearly all the age groups as well as E28 'Counter Cultural Mix'. The characteristics of this cluster are not necessarily unique and can be applicable to many other town centres, such as Camden late at night.

A6 'Pedestrians in the dark'

This hotspot type has 1968 casualties and 3937 drivers. It comprises of 4242 accidents and these are predominantly to pedestrians in the hours of darkness. This cluster has low proportions of under 16s and over 65s. It differs from A5 for two main reasons: first,, the collisions are likely to occur on main roads and, second, they are more likely to result in fatalities. Mosaic Types with the highest propensity to be involved in a collision as a casualty are E29 'City Adventurers', F36 'Metro Multiculture', E28 'Counter Cultural Mix' and E30 'New Urban Colonists'. These high propensity Mosaic Types indicate that Group E 'Urban Intelligence' have a high propensity to be casualties in road collisions late at night in built up urban areas within London. This Mosaic Type is likely to be involved in collisions along the Euston Road towards Kings Cross. In more detail, by profiling Mosaic Type E29, 'City Adventurers', it is apparent that there is a high proportion of young people aged in their early twenties being victims as pedestrians. Females rather than males from this Mosaic Type are more than three times as likely to be injured as a passenger. Overall the casualties in this Cluster exhibit an age pattern similar to the other Clusters, whereby there is a peak at the age of 32, with a steep rise from the age of 16, possibly indicating an overall increase in use of the road environment. Female casualties are twice as likely to be passenger casualties than males. For pedestrians the distribution of male and female casualties is fairly even, in comparison to drivers or riders who, like a large number of the Clusters are more likely to be male. With regards to Mosaic Type by age range, this Cluster is dominated in all age groups by E29 'City Adventurers'. The 65+ age group tends to have a slightly different over representation to the other age groups. For example casualties in this age group are three times as likely to be from Type E30 'New Urban Colonists' and they are also over represented in D21 'Respectable Rows'. The patterns of this cluster deem that lighting is not an issue as most of the hotspots occur along well lit main roads and that there is a possibility it may be pedestrians acting irresponsibly or without due care and attention in the road environment.

The drivers likely to be involved in this type of collision at certain locations are more likely to be from Type F36 'Metro Multiculture', E28 'Counter Cultural Mix' and E29 'City Adventurers'. The disaggregation of the driver dataset breaks down into 3001 male drivers and 765 female drivers, as well as 117 'unknown', where the sex of the driver has not been recorded by the police. Given the low average age of the Mosaic Type 'City Adventurers', it is not surprising that the peak age of drivers involved in collisions is aged 24-38. There are, as well as cars, a high number of motorcycles and pedal cycles involved in collisions. The drivers from this Cluster by age group are dominated by more Inner London Mosaic Types (unlike types such as H46 'White Van Culture' or C18 'Sprawling Subtopia').

A7 'Weekday rush hour pedestrian hotspots'

This Cluster is characterised by the increased number of pedestrians being injured during weekday rush hour traffic. The Cluster has a total of 2811 collisions scattered across London. Of these collisions there were 1760 casualties and 3361 drivers involved. Firstly the casualties, Mosaic Type D26 'South Asian Industry' had a high over representation amongst those injured, three times as likely with an index score of 327. This was followed by Mosaic Type F36 'Metro Multiculture', E28 'Counter Cultural Mix' and E29 'City Adventurers'. Therefore, overall the mix of people likely to be casualties is largely people from an ethnic minority, with a smaller proportion of younger more affluent people. The breakdown of casualty type by sex mirrors the pattern in many of the Cluster types, where there are largely similar numbers of pedestrians being injured, compared to a higher proportion of female passengers and male drivers. This Cluster's drivers and casualties disaggregated by age has a heavily over represented number of pedestrian children being injured aged 11-12 of ethnic minority origin. There is also an increased propensity to be involved in a collision amongst the 28-30 year old group.

The drivers of this Cluster are not only disproportionately drawn from types F36 'Metro Multiculture', D27 'Settled Minorities', and C20 'Sprawling Subtopia', but more significantly from Mosaic Types A01 'Global Connections' and A02 'Cultural Leadership'. In Chapter Four, we saw that both of these Mosaic Types were vastly under represented in both the casualty and driver categories. The drivers from these two Mosaic Types are more likely to be male, and be the driver of a vehicle. There is also a high propensity for the driver to be turning right at the time of the collision. As well as the high propensities for Mosaic Types A01 and A02, type C20 'Original Suburbs' is nearly twice as likely to be involved in a collision as a driver. The under 15 drivers (or possibly cyclists) are dominated by ethnically diverse Mosaic Types including especially D26 'South Asian Industry' and F36 'Metro Multiculture'. The 15-19 age group has over exposure in D27 'Settled Minorities', F36 'White Van Culture' and E28 'Counter Cultural Mix'. This pattern continues up to the 55+ age groups. The 55-64 and 65+ age groups are more likely to be from E30 'New Urban Colonists', B12 'Middle Rung Families' and C15 'Close to Retirement'.

A8 'Morning commuting cyclists at rush hour'

This Cluster is characterised by collisions involving cyclists in the morning rush hour. The Cluster has 7720 collisions, with 4885 casualties and 9943 total drivers involved. Cluster A8 has an over representation of Mosaic Types F36 'Metro Multiculture', E28 'Counter Cultural Mix', and D26 'South Asian Industry'. The percentage of casualties under 16 and over 65 is low. This indicates that this Cluster is not made up of particularly vulnerable road users (such as children and the elderly). The dataset reveals that of the children under 16, a large proportion are passenger casualties, they are more likely to be in a car and more likely to come from F36 'Metro Multiculture' or E28 'Counter Cultural Mix'. There is an increased number of drivers and riders in their twenties which would reinforce the pattern of the Cluster having a high number of cyclist casualties. These drivers in their 20s are more likely to be from either F36 'Metro Multiculture' or D26 'South Asian Industry'.

The drivers of this Cluster show very different characteristics compared to those involving casualties. The Cluster has a large number of over represented Mosaic Types for all the top 14 Mosaic Types which make up a large proportion of London's population. This includes Type A01 'Global Connections' and J52 'Childfree Serenity'. This large number of over represented Mosaic

Types indicates a pattern of drivers who are more likely to come from a wide range of Mosaic Types. Therefore it is impossible to pinpoint a direct and clear pattern of driver involvement. The predominance of a Central London location for this Cluster suggests that the casualties who are predominantly cyclists would come from nearby whereas the drivers would be from more diverse and further a field locations based on their Mosaic Types. Each of these Mosaic Types would be at risk from being a driver hitting a cyclist in the Central London morning rush hour. Supporting this, is the under representation of H46 'White Van Culture' in this Cluster for drivers. The older drivers in this Cluster (55+) are more likely to be from A01 'Global Connections' and A03' Corporate Chieftains'.

B9 'High density careless weekend drivers'

This Cluster has a small number of collisions compared to the other Clusters, with only 584 collisions. With its 265 casualties and 479 drivers involved in collisions within this Cluster, this is one of the smallest of the 15. The index scores for this Cluster for the most at risk Mosaic Types are dominated by type E28 'Counter Cultural Mix', E30 'New Urban Colonists', E29 'City Adventurers' and A01 'Global Connections'. There is a low proportion of children involved in collisions in this Cluster but the highest number of over 65s. These older casualties are more likely to originate from two different socio economic types: E28 'Counter Cultural Mix' and also A01 'Global Connections' and A02 'Cultural Leadership'. This Cluster is characterised by involvement of weekend drivers who are more likely to be older (over 65) This Cluster has nearly three times as many drivers as pedestrians and passengers. There is a high number of over 65 year old women who are passenger casualties within this Group.

The drivers who are more likely to be involved in these high density weekend collisions are not known. The propensities for the Mosaic types for this Cluster are all under represented, therefore indicating that there are no obvious associations between drivers and Mosaic Types within this Cluster. There are no Mosaic Types which have an over represented propensity to be involved in a collision.

C10 'Cyclists in Westminster'

This is one of the most distinctive Cluster types compared to the other fifteen Clusters, as it covers a large and unique area of central London including Oxford Street, Edgware Road, Charing Cross road and Soho. It has a total of 7132 collisions occurring in it, and involving 3306 casualties and 5930 drivers. Of the casualties only 4% were aged 16 and under, indicating a low proportion of children involved in collisions in this area -- but this is to be expected given the area of London. However there was a slightly higher percentage of over 65s injured in these areas. These over 65s are likely to come from a diverse background: A01 'Global Connections' is over represented in this age group as is E28 'Counter Cultural Mix' and B12 'Middle Rung Families'. This Cluster has four Mosaic Types which are over represented, E28 'Counter Cultural Mix', E29 'City Adventurers', F36 'Metro Multiculture' and A01 'Global Connections'. This highlights a cross section of society which is involved in collisions, from the very rich and highly successful who live in central London to those who are less affluent and who live further out of London. There are 1337 male drivers/riders, compared to only 295 females. In comparison there are 471 female passengers and 296 male passengers. The counts for pedestrians are approximately equal between males and females. The vulnerable road user in this Cluster is clearly cyclists and pedestrians. With regards to age breakdown in relation to Mosaic Type in this Cluster for casualties, E29 'City Adventurers' dominate the 20-44 age range, whereas A01 'Global Connections' is over represented amongst the older age ranges.

The pattern for high risk drivers is very different. Nearly every Mosaic Type which constitutes over 1% of the London population is over represented. Some of the highest index scores belong to Mosaic Types such as C19 'Original Suburbs', and C18 'Sprawling Subtopia' possibly indicating that people are driving into London from the suburbs of London and hitting cyclists and pedestrians which are more central London based. The vehicle type is obviously dominated by cars, however there are high numbers of both motorcycles and pedal cycles, closely followed by buses (and coaches) and taxis. This pattern indicates that there is a strong pattern of bus, taxi and car involvement in collisions with two wheeled vehicles. Due to the clusters central location there are high numbers of buses and taxis in this area of London. The younger drivers involved in this Cluster are more likely to come from a polarised background, with fewer Mosaic Types being over represented in the under 15 and 15-19 age groups. E28 'City Adventurers' are over represented throughout all the age groups in this Cluster, compared to, for example, D21 'Respectable Rows' which is over represented only in the 15-19 age group.

D11 'Dual Carriageway joy riders'

Compared to the other Clusters, D11 has 1828 collisions, which is small. This Cluster has 1390 casualties and a total of 2588 drivers, which is just less than double the number of casualties. Disaggregation of the casualties reveals that 7.8% of them are aged 16 and under: compared to the other Clusters this is just above average. The index scores for this Cluster reveal that F36

'Metro Multiculture' has the highest over representation. Of this high propensity group, the percentage of casualties 16 and under is 21.9% which is very high. Unlike many of the previous Clusters, D11 has 93.4% of its F36 'Metro Multiculture' casualties which are predominantly drivers or passengers. The drivers are predominantly male, whereas the passengers are more likely to be female. Other Mosaic Types which have high index scores are D27 'Settled Minorities', C19 'Original Suburbs' and B12 'Middle Rung Families'. Casualties aged 55-64 in this Cluster are more likely to come from either C15 'Close to Retirement' or A03 'Corporate Chieftains'. The 25-44 age range is dominated by an over representation of E29 'City Adventurers'. The 15-19 year old casualties are more likely to be from B12 'Middle Rung Families' or D27 'Settled Minorities'. In this Cluster the older age ranges tend to experience a higher polarisation in terms of Mosaic Types compared to the younger age ranges.

The drivers in this Cluster have a very different pattern to the casualty index scores. There are low index scores for F36 'Metro Multiculture' and D27 'Settled Minorities'. The high risk exposure Mosaic Types include A01 'Global Connections' and H46 'White Van Culture'. This cluster is dominated strongly by cars, with a percentage of 74.3%, highlighting the high occurrence of car on car collisions. Whilst the overall index scores for this Cluster reveal under representation the index scores by age range reveal slightly different patterns. The under 15 drivers are likely to be from either B12 'Middle Rung Families' or C19 'Original Suburbs'. The former Mosaic Type is also over represented in the 15-19 age range as well as Type A03 'Corporate Chieftains' and C20 'Asian Enterprise'. The older age ranges are dominated by C15 'Close to Retirement' and C18 'Sprawling Subtopia'.

D12 'Main road multiple victim collision in Outer London'

This Cluster is characterised by many casualties on main roads in Outer London. It has 2974 collisions which are made up of 2308 casualties and 4156 drivers. The percentage of casualties in D12 which are aged 16 and under is high. The 20-24 age group is more likely to be from C20 'Asian Enterprise' and C15 'Close to Retirement' which seems a potential anomaly in the results. In this Cluster E29 'City Adventurers' (especially in the 20-44 age group) are under represented, and there are over representations of H46 'White van Culture' in the middle age ranges. The older age ranges especially the over 65s are dominated by D21 'Respectable Rows' and C15 'Close to Retirement'. The Mosaic Types which have an over representation is H46 'White Van Culture', C18 'Sprawling Subtopia', C15 'Close of Retirement' and B12 'Middle Rung Families'. In Mosaic Type H46 'White Van Culture' 17.5% of casualties are aged 16 and under. There is a small proportion which is aged 65 and over. Within this type there are 60% of the casualties which are drivers, 36% which are passengers and 3% which are pedestrians. This highlights the Outer London, main road phenomenon as there are fewer vulnerable people being injured in terms of pedestrians.

Drivers in this Type are more likely to come from a large number of different Mosaic Types including F36 'Metro Multiculture' and E29 'City Adventurers'. Overall though, within this Cluster, the drivers are more diverse and have a less polarised propensity than the casualties. The prominence of the car being involved in 76.6% of collisions indicated that this Cluster is dominated by vehicles colliding with each other or with road furniture rather than hitting pedestrians or cyclists. Drivers by age group portray some differences compared to the casualty age indexes. For example A03 'Corporate Chieftains' have an over representation in nearly all the age ranges, apart from 25-34 which is dominated by H46 'White Van Culture'.

D13 'Sunday afternoon multiple casualties'

This collision Cluster is characterised by occurrences on Sunday afternoons that cause multiple casualties. There are 2706 which occur within this Cluster, including 2420 casualties and 4277 drivers. The over represented 'high risk' Mosaic Types in this Cluster are type D27 'Settled Minorities', C20 'Asian Enterprise', H46 'White Van Culture' and C18 'Sprawling Subtopia'. Over 95% of the total casualties within this Cluster are either drivers or passengers involved in collisions, with an equal proportion of both males and females. This is unique in relation to the results from the other Cluster types. Children aged 16 or under in this Cluster are more likely to be pedestrians rather than cyclists. The younger casualties are likely to be from H46 'White Van Culture'. As the age increases in this Cluster, the casualties are less likely to be from certain Mosaic types, in other words the likelihood of being involved in a collision, on average decreases with age. The 25-34 age range is dominated by E32 'Dinky Developments' who are three times as likely to be involved in a collision in this Cluster.

The pattern of driver geodemographics in this Cluster mirrors a similar pattern to the other Cluster types insofar as there are a high number of Mosaic Types which have an over representation of being involved as a driver. Both Mosaic types E28 'Counter Cultural Mix' and E29 'City Adventurers' have Mosaic index scores of over 400. The vehicle type is dominated by the car, and there are only low numbers of pedal cycles, taxis and buses involved. The under 15 drivers/cyclists are more likely to be from D26 'South Asian Industry', B12 'Middle Rung

Families' and C20 'Asian Enterprise'. In the 15-19 age group there is a higher index score in C20 'Asian Enterprise' for drivers rather than casualties. The 20-24 age group is dominated by D26 'South Asian Industry' and C20 'Asian Enterprise'. Both of these have strong ethnic characteristics and depict a heightened risk within this Cluster of being involved as a driver. The over 65 casualty in this Cluster has a high propensity to derive from either D21 'Respectable Rows' or B12 'Middle Rung Families'.

D14 'Risk taking early risers'

This Cluster is characterised by early morning commuter collisions frequently at large junctions in both Inner and Outer London. The Cluster comprises of 2613 collisions, involving 1981 casualties and 3774 drivers. The Cluster which is in collision Group D has similar casualty index scores. For example Type D27 'Settled Minorities' have a general over representation in this group as well as H46 'White Van Culture' and C18 'Sprawling Subtopia'. C20 'Asian Enterprise' has only a very small percentage of children aged 16 and under involved in collisions, indicating a predominance of drivers involved rather than young pedestrians of cyclists. There is also a high percentage of drivers (70%) and only 30% are passengers – a low figure compared to the other Clusters. Therefore it seems that the majority of casualties are drivers rather than passengers or vulnerable road users such as pedestrians or cyclists.

The people who are more likely to be a driver in a collision such as the one described above, are, like many of the other Clusters likely to be diverse in so far as there are a number of Mosaic types which are over represented. Similar to D12 and D13, the high index scores are prevalent in all three collision types. By far, drivers from type A01 'Global Connections' are more likely to be involved as a driver who is also a casualty. Other Mosaic Types which have an over representation are E28 'Counter Cultural Mix', E29 'City Adventurers' and E30 'New Urban Colonists'. Group E includes large proportions of the Mosaic Types which have above average propensities to be involved in a collision. This pattern occurs throughout the Group D 'collision types', indicating that people who live in areas in Group E are more likely to be involved in collisions in Collision Group D. The driver age for this Type peaks at 30. These 25-34 aged drivers are more likely to be from H46 'White Van Culture'. All of the age groups are dominated by this Mosaic Type particularly the older age groups. Unlike many of the other Clusters which are predominantly focused in Central London, Cluster D14 has a very small percentage of pedal cycles and pedestrians involved, indicating these collisions and people involved in them are less likely to reside in Central London.

E15 'Sunday morning pedestrian risk takers'

This Cluster only consists of two hotspots and is characterised as early Sunday morning pedestrian risk takers. In total, this collision type has 82 collisions involving 89 casualties and 171 drivers. E15 has over representations from type D27 'Settled Minorities', C20 'Asian Enterprise' and D21 'Respectable Rows'. E15 has no pedestrians involved, indicating that all the casualties were drivers (possibly of pedal cycles) and passengers. The under 15s involved in this Cluster are likely to come from either C20 'Asian Enterprise' or D21 'Respectable Rows' or D27 'Settled Minorities'. The older over 65s have a propensity to be from J52 'Childfree Serenity', and the pattern for drivers in this Cluster remains similar to that of the drivers Mosaic Type propensities.

The driver index scores are all under represented, making it difficult to draw any geodemographic conclusions about the types of people more likely to be a driver in this type of collision. There are only two pedal cyclists involved in collisions, with 87.7% of the vehicle type belonging to cars. Overall it is impossible to make any accurate conclusions based on the small numbers of casualties and drivers.

Cluster No.	A1	A2	A3	A4	A5	A6	A7	A8	B9	C10	D11	D12	D13	D14	E15
Total. Casualties	9	2046	4	2215	3316	1968	2811	4885	265	3306	1390	2308	2420	1981	89
Under15%	11.11	9.73	0.00	6.91	5.88	5.84	3.77	4.85	3.77	2.93	6.19	7.58	6.20	5.50	3.37
15-19%	0.00	9.82	0.00	6.14	5.88	7.52	3.84	5.22	6.42	4.90	8.13	7.37	9.75	8.78	4.49
20-24%	0.00	11.63	0.00	10.84	11.34	12.30	8.08	11.55	9.43	13.67	14.46	13.86	13.14	12.42	11.24
25-34%	22.22	23.12	50.00	29.57	31.60	35.37	20.53	34.33	32.45	34.21	30.29	29.98	27.56	29.03	28.09
35-44%	11.11	17.94	0.00	20.05	21.08	19.26	12.49	22.54	20.00	18.91	19.93	18.11	19.01	20.90	26.97
45-54%	0.00	9.24	0.00	10.20	9.11	7.57	6.01	8.76	6.42	9.41	8.27	9.71	9.42	9.74	15.73
55-64%	33.33	5.52	25.00	5.06	5.01	4.67	2.92	4.89	8.30	5.75	4.39	4.25	5.41	4.54	5.62
65+%	11.11	8.75	25.00	6.50	5.61	3.25	2.53	3.89	7.17	5.81	3.38	4.51	3.47	3.74	2.25

Table 7.3: *Percentages of casualty age groups within each Cluster*

In Table 7.3 we can see that A1 is the smallest Cluster and therefore the percentages are fairly large, indicating that most of the casualties are aged 55-64. It is clear at this stage that the largest age group is between ages 25-34. This age group has the largest proportion of casualty involvements. However the importance of this table enables the reader to depict which Clusters have higher or lower percentages. For example A2 has a high percentage of casualties under 15 involved in the collision, whereas A3 and C10 have very few. Group D and its associated Clusters has large percentages of casualties aged 15-19. This pattern also exists for the 20-24 age group. However the 25-34 age group has a more varied pattern, whereby there are very high percentages in Cluster A6 and C10 -- predominantly central London Clusters rather than those in the suburbs. The 35-44 age range has a high proportion of people in E15 being involved, as well as A4 and A5. From the age of 45 upwards the involvement in collisions in each Cluster decreases dramatically. This could be for two reasons: either they are safer on the roads, or their presence in the road environment is proportionally less than the younger age groups.

Mosaic Type	London	London %	Casualty %	A1	A2	A3	A4	A5	A6	A7	A8	B9	C10	D11	D12	D13	D14	E15
D27 Settled Minorities	843155	11.38	12.7	110	154	220	154	122	100	104	107	97	78	139	114	140	127	118
F36 Metro Multiculture	912740	12.32	12.3	203	111	203	169	193	171	165	184	102	161	169	103	81	98	27
E28 Counter Cultural Mix	675251	9.12	8.2	0	135	0	137	194	145	148	162	230	189	101	81	48	45	37
C20 Asian Enterprise	478737	6.46	6.7	193	93	0	105	60	77	65	104	47	59	95	109	179	188	174
H46 White Van Culture	348466	4.70	6.2	0	122	0	113	48	60	68	51	49	42	21	171	228	209	119
C19 Original Suburbs	467482	6.31	6	0	108	0	71	59	54	57	55	42	51	126	91	111	71	71
E30 New Urban Colonists	492125	6.64	5.1	0	105	376	78	110	140	113	102	172	98	100	82	42	57	0
C18 Sprawling Subtopia	273904	3.70	4.1	0	59	0	55	44	47	75	67	83	43	58	135	135	139	365
E29 City Adventurers	432742	5.84	3.9	0	87	0	57	124	217	120	110	144	164	107	56	34	32	38
D21 Respectable Rows	188055	2.54	3	0	108	0	58	53	75	87	56	15	52	72	104	107	107	487
A02 Cultural Leadership	344224	4.65	2.9	0	45	0	54	73	46	54	58	90	79	87	85	54	60	0
A01 Global Connections	393595	5.31	2.2	0	65	0	30	51	54	63	66	115	154	33	59	35	8	63
A03 Corporate Chieftains	164708	2.22	1.9	0	45	0	37	29	19	31	62	34	93	68	87	56	39	51
C15 Close to Retirement	88796	1.20	1.7	0	42	0	50	38	34	73	73	127	51	89	169	143	115	94
B12 Middle Rung Families	85951	1.16	1.7	0	77	0	79	66	58	65	65	0	53	124	167	209	185	290
J52 Childfree Serenity	147971	2.00	1.6	0	90	0	85	60	44	79	42	57	54	99	68	34	54	112
E32 Dinky Developments	92444	1.25	1.6	0	132	0	44	54	83	65	79	61	89	91	113	168	148	270
A05 Provincial Privilege	141025	1.90	1.5	0	39	0	63	23	41	49	49	40	29	52	69	101	62	0
D26 South Asian Industry	85644	1.16	1.2	3243	125	0	119	64	143	327	132	66	75	85	72	163	155	0
H47 New Town Materialism	48872	0.66	1.1	0	61	0	167	65	23	123	28	0	75	0	80	114	248	0

Table 7.4: *Index scores for casualties in all Clusters sorted by % of casualties and then % of London population*

Mosaic Type	London	London %	Driver %	A1	A2	A3	A4	A5	A6	A7	A8	B9	C10	D11	D12	D13	D14	E15
D27 Settled Minorities	843155	11.38	12.3	329	147	220	149	124	109	189	508	31	476	108	227	194	175	9
F36 Metro Multiculture	912740	12.32	11.4	152	110	203	139	170	163	128	320	32	249	96	272	363	245	40
E28 Counter Cultural Mix	675251	9.12	7.5	0	128	0	130	167	147	106	269	10	157	119	257	456	396	22
C20 Asian Enterprise	478737	6.46	7	0	98	0	121	75	76	172	297	36	357	89	135	86	67	3
C19 Original Suburbs	467482	6.31	6.3	99	106	0	72	62	56	190	548	39	402	66	158	136	174	8
H46 White Van Culture	348466	4.70	6.1	0	106	0	120	68	67	119	446	25	363	295	62	49	44	4
E30 New Urban Colonists	492125	6.64	4.9	0	108	376	91	104	123	101	312	10	219	88	184	376	227	0
C18 Sprawling Subtopia	273904	3.70	4.3	0	66	0	75	55	57	84	266	12	276	85	62	66	52	1
E29 City Adventurers	432742	5.84	3.7	0	83	0	65	128	174	84	255	11	115	72	238	414	351	14
D21 Respectable Rows	188055	2.54	3.1	0	89	0	88	66	81	50	218	44	157	47	55	57	46	0
A02 Cultural Leadership	344224	4.65	3	0	58	0	54	60	54	149	382	13	191	71	124	205	152	0
A03 Corporate Chieftains	164708	2.22	2.3	0	45	0	66	48	33	122	172	17	78	43	58	94	111	4
A01 Global Connections	393595	5.31	2.1	0	53	0	33	57	56	144	384	12	111	213	205	366	1345	7
C15 Close to Retirement	88796	1.20	1.8	0	61	0	71	60	61	28	79	2	75	18	16	20	20	1
B12 Middle Rung Families	85951	1.16	1.8	0	94	0	77	77	83	31	86	0	71	12	16	13	12	0
J52 Childfree Serenity	147971	2.00	1.6	0	101	0	61	50	39	44	229	9	120	27	66	142	72	2
A05 Provincial Privilege	141025	1.90	1.6	0	38	0	55	37	38	67	185	12	212	48	62	45	60	0
E32 Dinky Developments	92444	1.25	1.6	0	166	0	87	75	108	33	76	5	45	18	25	18	16	0
D26 South Asian Industry	85644	1.16	1.2	3243	105	0	134	65	177	6	42	5	50	18	36	17	15	0
H47 New Town Materialism	48872	0.66	1.1	0	115	0	169	61	75	9	111	0	29	0	19	14	5	0
A06 High Technologists	42654	0.58	1	0	95	0	130	58	45	33	54	2	39	57	19	15	7	0
B11 Families Making Good	37249	0.50	1	0	163	0	67	63	88	8	22	0	17	5	12	6	7	0
B13 Burdened Optimists	31072	0.42	1	0	196	0	87	123	87	7	14	0	15	3	5	5	3	0

Table 7.5: *Index scores for drivers in all Clusters sorted by % of drivers and then by % in London (All drivers over 1% selected)*

7.6 Example of how typology could be utilised

This section teases out how this typology may be used in a 'real world' scenario. It discusses two different examples; the first is from the perspective of using a residential location (a postcode) to disseminate someone's risk or likelihood of being involved in a certain collision. The second example describes how one would predict the type of person involved in a specific collision (when knowing the collision location (see Figures 7.2 and 7.3). Method One begins with the knowledge of an individual's postcode, with the aim to understand that person's collision risk based on their surrounding area in which they live. Using a London wide postcode analysis, one can determine which Mosaic Type the postcode falls into. Using this Mosaic Type it is possible to append the risk 'likelihoods' of being a driver or casualty in a collision. It is also possible to deem the type of collision more likely to be involved in.

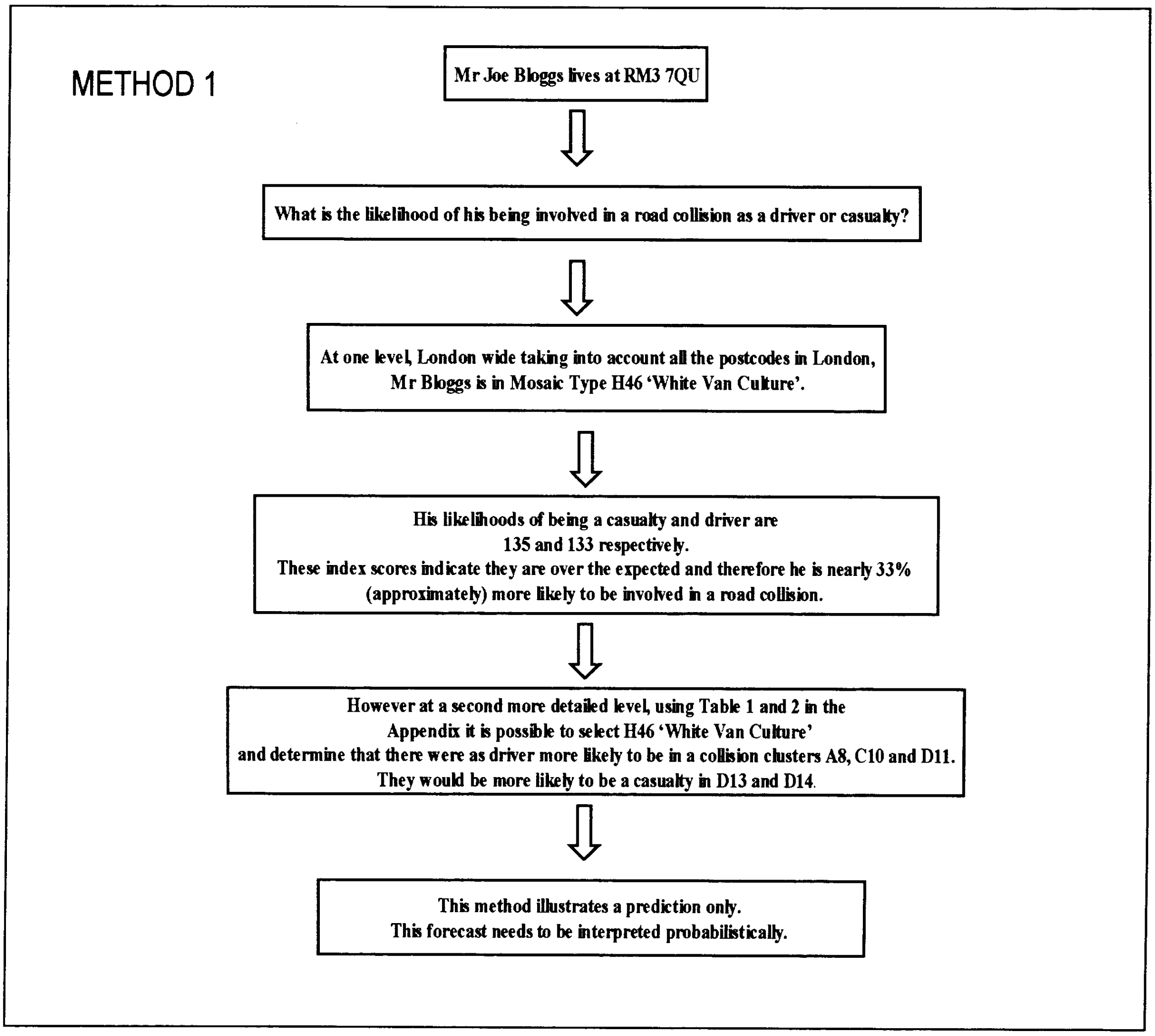


Figure 7.2: *Method One – Using a person's postcode to predict collision location involvement*

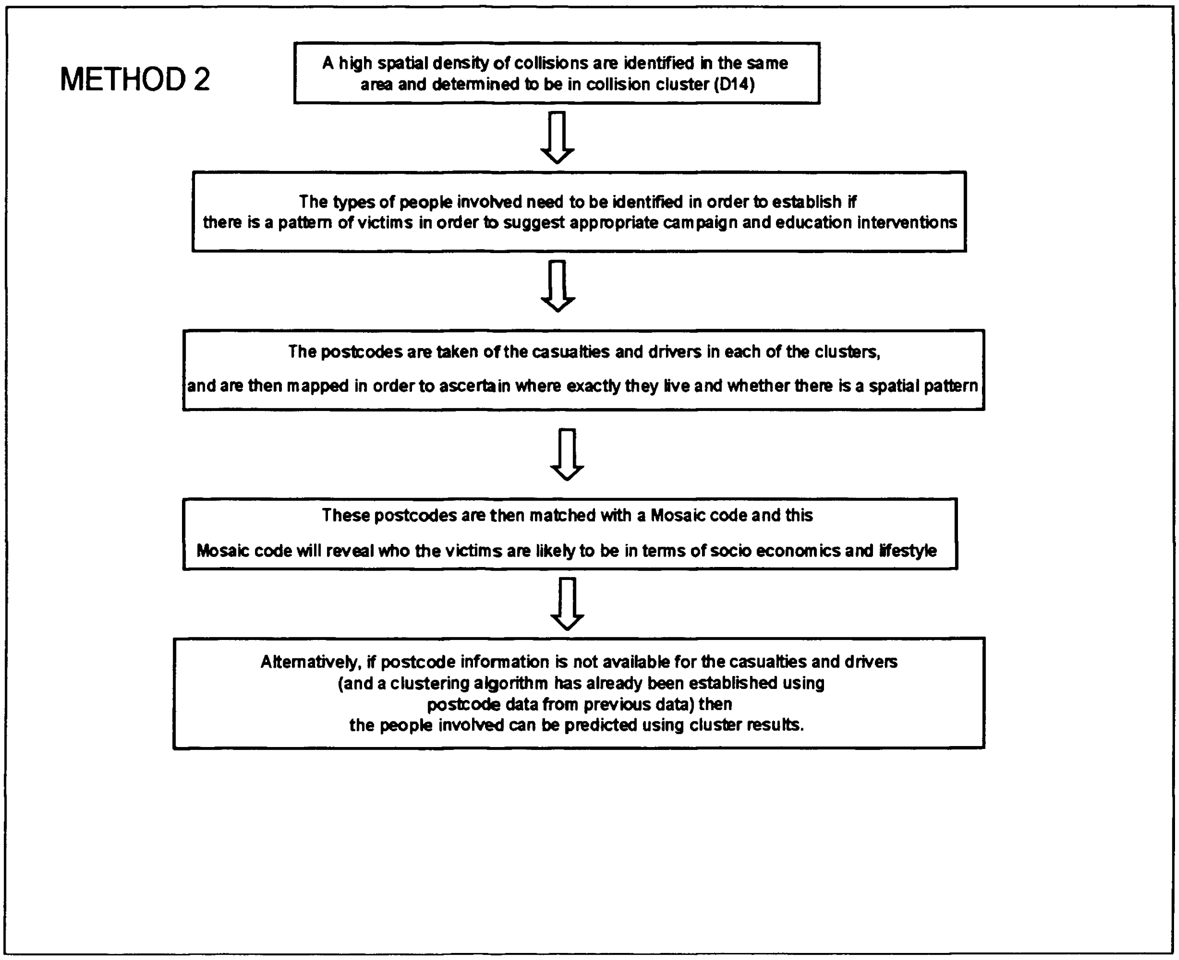


Figure 7.3: *Method Two- Using a collision location to predict the type of person likely to be involved*

Figure 7.3 shows the utilisation of the typology from the other geographical perspective, and this is the methods which have been used in this thesis. It begins by determining a road collision cluster and it is deemed necessary to identify the people involved in the hotspot in order to establish any potential patterns and for appropriate communication campaigns. The postcodes of the drivers and casualties involved in the hotspot are taken and mapped. These postcodes are then appended to a Mosaic Type and will reveal the 'types' of people who are likely to be involved in the collisions based on the surrounding location of where they live.

7.7 Profiling London's population: road collision exposure based on socio economics and lifestyle characteristics

It is all very well outlining the Mosaic Types and Groups which are more likely to be involved in road collision clusters however it is important to step back, and outline what this means in a broader sense and to try and unpick the patterns and processes which are going on within the collision clusters and groups. By ascertaining the types of people who are more likely to be exposed to certain collision types, we can gain a better interpretation of what is really happening. It also highlights perhaps a number of neglected and under researched areas which are clearly affecting road safety in London. The themes which are present in this typology include; the strong ethnic minority trends, attitudes to cycling, possible passenger distraction, young adults, irresponsible pedestrians and inner and outer London differences. Each of these themes will be addressed in turn.

One of the most prominent themes which is apparent throughout all of the clusters is the strong ethnic minority trends. Nearly all of the clusters have over represented index scores for F36 'Metro Multiculture', E28 'Counter Cultural Mix' and D26 'South Asian Industry'. These Mosaic Types consist of a large proportion of ethnic minorities, and indicate that they have a higher than average collision risk exposure. In Chapter Four, the relationship between road collision and ethnic minority children was investigated and a relationship was found: however results from Chapter Four and Seven subsequently suggest that adults from minority populations are also at risk from a higher exposure to road collisions. Immigration is an important social issue in London because of the high numbers of immigrants to the capital. Social issues surrounding health, welfare and employment have been addressed; however there has been no research on the propensity of the adult population from minority populations to be involved in collisions. Transport for London acknowledges that boys aged 11-14 from low income and ethnic minority backgrounds are particularly vulnerable to road collisions. However there is no evidence that the same is true for adults from ethnic minority backgrounds. This thesis suggests that research is needed in this area in order to effectively target the appropriate population groups. Transport for London's road safety plan (2001) does not recognise the social trends of road collision involvement in London, and instead focuses upon pedestrians and cyclists as vulnerable road users, which seems obvious when they are always going to be at the mercy of vehicles. One of the central aims of this thesis has been to suggest ways of breaking up these heterogeneous groups.

The second theme which emerges from this typology outlined in this chapter is the noticeable risks of cycling in central London. The casualties of these collisions are from a diverse background including A01 'Global Connections' and F36 'Metro Multiculture'. Although it is common knowledge that cyclists are vulnerable road users, these results show us a little more information. Firstly the unique spatial location (with regards to the fact that it covers a large proportion of Central London), indicates there is larger problem in Central London. Although Mayor Ken Livingstone has vowed to make London into a 'cycling city' the problem lies in the interaction of cyclists in the busiest traffic area of London. There are no clear cycle lanes and no law about wearing cycle helmets. It clearly shows that something needs to be done in this area, not only tackling the environment but targeting the types of road users who are the casualties of these collisions.

The third theme which seems to be emerge from the typology is the role of the passenger. A large proportion of all the collisions in every cluster of the typology involves vehicle passengers. This, although we cannot be certain begs the question of passenger distraction. This occurs in the clusters in Group D which is characterised by its Outer London location and people from an ethnic minority background. Very little research has been conducted in this area because of the subjective nature and difficult nature of data collection, there is no way of determining whether passengers distracting the driver caused the collision. However it is an obvious problem especially amongst the younger generations of passengers causing the driver to lose concentration. This theme, as with the others warrants further investigation in order to ascertain this risk group.

In a similar perspective the next theme relates to the identification of young adults as a high risk group. There is so much focus on children in London, both as pedestrians and cyclists, yet there is rather less attention paid to the young adult drivers who make up a large proportion of the risk groups -- especially in Outer London where the roads are less built up and speeds can be increased. This group are often newly qualified and have tendencies to show off and drive around with other young adults as passengers. Their risk taking behaviour endangers other road users and as with the other themes needs further research, specifically relating to the socio economic characteristics of these risk taking drivers.

The penultimate theme relates again to a vulnerable road user, the pedestrian. The majority of the inner city clusters (all of Group A) have a strong element of pedestrian casualties involved.

However from the results there is two types of pedestrian casualty. First, the responsible pedestrian, who is at the mercy of the vehicle: these people tend to be vulnerable road users (cluster A2) and experience heightened risk. However the second type of pedestrians who are often injured are the irresponsible pedestrian found in A5 and A6. These pedestrians often fail to interact with the road environment in a responsible manner. These pedestrians tend to be from E29 'City Adventurers' especially in A6 whereas the pedestrians in A5 tend to be more ethnically diverse.

The final theme which runs throughout all the clusters is the strong differentiation between Inner and Outer London and the prominent differences between the types of collisions and the casualties and drivers that are involved. For example Collision Group A has collisions which largely occur within Central London and the types of casualty and driver vary considerably compared to say for example Collision Group D. Group A has a high number of pedestrian and cyclists injured in the clusters, with drivers coming from all parts of London and very different Mosaic Types. By contrast in Group D, the collisions tend to be on major roads and there are fewer pedestrians and casualties involved. In each borough, the road safety reduction team often cannot compare types of road collision to other boroughs, although it is clear that collision clusters occur in different boroughs with the potential of the management of the collision cluster being the same or similar. This idea of cross-border partnership was seen as important in London's road safety plan of 2001.

7.8 Perspectives on road collision typologies

This chapter has endeavoured to unite the two main elements running through this thesis; the collision location and the home location of the driver and casualty. Planning, creating and analysing this typology has contributed to understanding of the patterns and processes which are working within the collision Clusters and Groups. Massie *et al* (1993) provided one of the first research papers to introduce this idea of a typology into road safety and road collision analysis. Even in 1993, there had been very little research in this area and Massie *et al*'s review of literature relevant to collision categorization found relatively little work in the area. Over the years, however, different scales and meanings have been introduced to the idea of classifying collisions. At a basic level, collisions have been classified as attributable to human, environment and vehicle factors in terms of collision error. Human error is implicated in 88-95% of all collisions in these studies (Perchonok 1972, Treat *et al* 1979, Sabey *et al* 1980). However, as the road safety literature has matured, so there has developed increasing acceptance that the causes of collisions are far more complex than once thought. Therefore the aims of this chapter have been to create a multi faceted typology which encompasses a range of potential risks and causes by focusing on the road environment and the socio economics of the people involved.

Because of the hitherto limited amount of research into creating collision typologies, the work in this thesis, and in particular in this chapter, has largely been exploratory. Therefore the resulting typology outlined in this chapter should be seen as an initial attempt and further analysis would need to track changes over time, preferably every year, in order to maintain an up-to-date pattern of risk involvement. Such analysis should be repeated frequently in order to provide greater accuracy and in-depth insight into the changing risk patterns in London. The limited time scale of this study made this impossible. The patterns within the Clusters revealed a strong Inner versus Outer London pattern, indicating that the people who are likely to be involved in a more suburban or Outer London; more car dominated collision the more likely the victim will be from the surrounding or nearby area.

This chapter therefore has detailed each Cluster and its corresponding characteristics of its casualties and drivers. This provides relevant and useful information for policy makers and road safety professionals alike. The index scores indicate a person's likelihood of being involved in a collision and the level of risk. The level of complexity surrounding the depiction and unravelling of collision events means that there is no universal benchmarking and that ultimately it is impossible to contain every single risk factor within a collision event as many are invisible to the naked eye, and are therefore not included. The typology in this thesis has tried to develop a unique and novel investigation to determine patterns of involvement to certain types of collision, clustered together in space.

CHAPTER 8

CONCLUSION

8.1 Summary of research findings

The aim of this thesis has been three fold. The first objective has been to use geodemographic analysis to ascertain the socio economic patterning of road collision involvement in London. The second objective has been to investigate the most appropriate method for road collision hotspot detection. The third objective has been to combine these two methodologies to create a road collision typology which accommodates residential location to predict when and where certain people are likely to have a road collision and vice- versa. Road collisions can have complex and multi-faceted causes, and road collision analysis may pose more questions that is answers.

The underlying question is whether it is possible to establish a relationship between a road collision outcome and causal socioeconomic and lifestyle factors. Chapter Four provides qualified evidence that this is an attainable goal. However, a further message from this chapter is that measuring risk exposure is hazardous and error prone. The solution adopted in this thesis has been to use the variation above or below expected values in order to ascertain whether a person is more likely to be involved in a collision. Risk in this instance is measured in terms of socio economics and lifestyle, and it is assumed that any individual's risk of collision will vary on account of this attribute. In this sense the concept of causation or contributory factor means the conditions which will alter the risk of a collision. Therefore the contributory factor is the socio economic characteristics of the driver or casualty involved in the collision. This contributory factor in this sense it is not spatially linked to the road collision site itself. The socio economic circumstances of an individual can greatly influence their exposure to a road collision and therefore indicate the level of risk to which they are exposed. The aim has been to measure this level of risk or exposure to road collisions. We have seen in the literature that there documented acute risk between young children, who are deprived and being more likely to be involved in a road collision. Other studies (Abdalla *et al* 1997) have concentrated on deprivation in particular rather than socio economics in general. This thesis has attempted to combine a large proportion of socio economic indicators to apply to the total population. In some ways this has made the task of this thesis harder because it has been necessary to consider every group within society. Many studies tend to focus on social groups such as the young, the elderly or perhaps for example motorcycle users which was discussed in depth in Chapter Three.

The spatial results of the likelihood to be involved in a collision in Chapter Four indicate patterns across space which can predict where people will come from and provide the basis for further analysis. In a sense the London wide analysis meant the risk and hotspot analysis concentrated on the total resident population of London. Further work would be beneficial at neighbourhood scale. This chapter had the benefits of providing information on, for example, the Mosaic Types such as H46 'White Van Culture' or D21 'Respectable Rows', which can provide further insight into the type of people they are. Enabling more detailed questions to be asked, for example, if a person is risky in their hobbies and with their money does this make them more risky drivers? It is also important to consider that Mosaic a classification product is designed to measure consumer behaviour for private sector applications. Therefore is it right to use to ascribe certain behaviours to road traffic collisions? This is an important issue which is not just an issue for this thesis but also other public sector applications using geodemographics such as health, crime and education.

This comes as road safety reduction measures reaches a stagnation point in terms of the number of engineering measures which can be implemented. A consequence of this may be that we are entering a new phase of road safety analysis which concentrates on educating the people of the risks of road travel. The only way to do this is to look at people's psychology and lifestyle and their relationship with the road environment. Linkage of STATS19 data to a casualty or driver's home address adds more value to the collision data and makes further analysis more robust. Furthermore, addition of social, economic and demographic information to the postcode within which the casualty or driver lives allows the user to make the association between collisions and social characteristics.

The corresponding spatial location is the collision location itself. This provides the basis for any road collision analysis and preventative measures. However in order to manage (reduce or prevent) the road collisions effectively there needs to be some measure of the 'worst' locations of collisions. Fundamentally this is the most contested area of road safety analysis. Chapter Five outlined the different methods which are associated with determining road collision hotspots. Ultimately from the diversity of literature surrounding collision hotspots it is apparent there is no right or wrong way of delineating hotspots. For this thesis, kernel density estimation was used to outline areas within London which had a high density of collisions. This was fundamentally because of the spatial nature of this method and how it rests of the spatial density of collisions. It provided a robust method for determining hotspots and enabled the user to make decisions regarding bandwidth and size of the basic spatial unit. This gave the user power over the results produced and aimed to create the most accurate representation of hotspots across London. There was a strong assumption that collisions occurring close to each other are exhibit spatial dependency and share a common cause. This assumption when analysing the clustering results in Chapter Six was deemed to be accurate and that hotspots not only had collisions which were similar but possessed a degree of similarity when compared to other hotspots (at a Group level). This clustering method aimed to create a more strategic method for the management of collision hotspots, by determining that within London similar hotspots of collisions would not occur independently and hotspot similarities would exist elsewhere in London. This clustering methodology provided the basis for understanding the relationship between causal factors within the hotspots. However the overall clusters have been based on a number of chosen indicators, for example the size of the basic hotspot was at $100m^2$ (the measure of resolution) (the average road length segment in London is 100m) and the number of clusters and groups which were predetermined. Had the number been different, would the characteristics of the clusters alter significantly? These are questions which will be discussed in the section 'Further research'. Ultimately the aim was to create an adaptable method that could be applied to different datasets easily. This method provides a visual representation of areas of high numbers of collisions. Its predominance in crime science is a testimony to its use. However one main drawback reoccurs, which relates to determining the statistical significance of the resulting clusters. This is an area of research which is something to investigate in further studies.

In an attempt to understand further the links and relationships between the two spatial locations, a typology was created by determining the Mosaic Types of the people involved in the collisions within the Collision Clusters and Groups. In an effort to disseminate who would be for example more likely to be involved in Cluster C10, covering a large proportion of Central London, the Mosaic Types were disaggregated by age in order to build a more accurate picture. It gives two methods of how one would determine patterns in road safety either from firstly knowing where a person lives and predicting the collision type they are more likely to be involved in. Secondly by knowing the high density location of collisions predicting who is more likely to be involved and where they are likely to reside. Typologies in general bear the scepticism of a wide range of people primarily because of the generalisations which are being made with regard to trying to assume a similar lifestyle for people who have the same postcode. Abdalla *et al* (1997) concluded that collision rates are higher among neighbourhoods with a high density population, smaller average household, high number of lone parents, more pensioners, fewer economically active adults, flats rather than semi/detached houses, lower car ownership, fewer people travelling to work by car, and higher percentages of migrant households. From this thesis it is clear that there is a relationship between neighbourhood and the likelihood of being involved in a road collision. Whilst Chapter Four provides a London wide analysis of who is more likely to be involved in collisions generally, the typology disaggregates these results by taking the highest density locations; therefore there are different types of people more likely to be involved in this level of road collision. In short we can summarise the two types of people likely to be involved in a collision: the first is a general pattern of involvement applied London wide with the ability to focus on specific neighbourhoods or areas such as Camden, Notting Hill, Tooting to determine the areas of residential addresses which are more at risk generally from being involved in a collision. The second type consists of identifying drivers and casualties likely to be involved in more high density collisions which have been identified using the clustering method. This utilises the method of prioritising areas where more attention is needed because of the scale of the collisions occurring. It should be emphasised, however, that these relationships are not necessarily causal and other, unquantified factors may be influencing the pattern of casualties. In particular, the relationship between casualty rate and neighbourhood will be affected by differences in exposure resulting from differences in travel pattern and mode of travel. Nevertheless, in drawing attention to the disparities in casualty rates between affluent and disadvantaged neighbourhoods, the findings of this research study provide useful information in assisting the targeting of local education and information campaigns, as well as informing road safety practitioners on the most vulnerable groups of road user and the area in which they live.

8.2 Policy recommendations and implications

This study aimed to gain a better understanding of the types of people who are more likely to be involved in certain collisions across London. By 'certain types', this was interpreted as people who had different socio economic lifestyles which was dependant on where they lived. Geodemographics was used to interpret if there were any patterns of involvement and socio economic 'type'. This meant that by establishing certain people more at risk of certain collision types, those people could be better informed with regards to road safety in their environment. The premise was that this education in road safety would possibly be based at a neighbourhood level across London. Arguably it is generally useful for identifying the best communications channels to contact different types of people however not all these channels will be local. There has been considerable growth of interest in neighbourhood issues including in particular crime, access to services and transport. There is a slowly developing interest in community road safety of both local neighbourhood scale and borough initiatives who have been taking a stake in trying to reduce the number of deaths on the roads. There have also been some communities that are beginning to address the road safety problems directly. The key in this instance is better intelligence so that neighbourhoods are more aware of certain road safety issues that may not be immediately obvious. The Department of Transport proposed guidelines at the beginning of 2003 for local authorities tackling road safety implications for disadvantaged areas. They outlined three areas of evidence for this. Firstly, that children from Social Class V are five times as likely to die as pedestrians than children from Social Class I. Secondly, it has been estimated that there would be 600 fewer road deaths amongst men aged 20-64 if all had the same accident probabilities as Social Class I. Thirdly, that road traffic danger and accidents can contribute to social isolation particularly among the elderly who have less access to cars. The objective is to reduce casualties in deprived areas (defined as the 10% most deprived wards as ranked by the Index of Multiple Deprivation: IMD 2000). However this thesis has outlined that although generally, deprived areas and residents are more at risk, this pattern is slightly more complex, particularly in London whereby the pattern of collision risk involved not necessary deprived residents but rather residents of non UK descent or those who were more likely to take advantage of the road environment. Generally speaking, this link between road collisions and deprivation in more complex and needs further work to disaggregate the different patterns of involvement.

Community and neighbourhood road safety initiatives therefore are the key policy recommendation for this thesis and the need for a wider collaborative body to initiate understanding, education and awareness at a local level within their environment. This means that community road safety groups have the potential to reach individuals and groups which are not reached by conventional media. Media campaigns aimed at raising awareness of issues or trying to influence road user behaviour are inevitably aimed at the demographic groups which form the core audience of these media. While some differentiation of audience is possible through radio, or magazines etc this communication can have drawbacks. Most obviously, people who do not understand English may well not receive the intended message, while other marginalised groups may reject messages aimed at the mainstream as not being relevant to them personally. Community road safety has the potential to overcome some of these problems by careful use of translated messages and by using persons to whom the marginalised groups can relate to deliver the messages. In a paper in 2000, Brown *et al* (2000) outlined this shift to targeting resources into small areas and how it is essentially used for the reducing crime, and addresses many community safety issues. However they argue that there needs to be some area-based priority setting. It outlines the use of geodemographics for assessing the priority of areas in order to identify the areas, before known courses of action can be taken in developing measures to reduce the rate of incidence.

The previous part of the policy recommendations has exclusively applied to the application of geodemographics to road collision victims in order to determine certain neighbourhoods and people more at risk. However, there is an important part of the thesis which also gives pause for thought. Most studies of GIS and road collisions have focused on the analysis of hotspots, whether this is finding them in the first instance of analysing them after this. This thesis aims to do both using kernel density estimation and a clustering algorithm. In Chapter 5, it was outlined that there has been a significant number of hotspot identification methods which vary on many different levels. This thesis provides a unique method for the interaction of both the hotspot location and the resident home location. The fundamental reason for this method is the notion of 'risk spreading' whereby is someone has a collision in a certain location, the likelihood of having a collision again in the same place is increased and also in the immediate surrounding areas. Kernel density estimation is proving a popular tool for analysis of road collision hotspots for the its main interpolation characteristic and being able to place values where they might not have been any road collisions, but acting more as possibly prediction tool for areas that might have collisions in the future. Chainey and Reid (2002) outlined that kernel density smoothing has become one of the more robust and appropriate techniques using for mapping patterns of crime. So therefore, as the aims of crime analysis mirror those of road safety, so kernel density smoothing should be seen as a robust and appropriate method for road safety analysis as well.

8.3 Future research

This section aims to provide recommendations and suggestions for possible priority directions for future research in the field of road collisions. Road safety strategy and funding rely heavily on research and understanding of the road environment. More often there is a shift from merely investigating the sites which have high collision rates to a more comprehensive method of investigating collisions. The many facets of this thesis have provided a much needed beginning for the next steps for road collision research. The typologies that have been created in this thesis provide a snapshot of the processes which are occurring at these sites and the people upon whom they impact. Perhaps inevitably, this warrants further site investigation with regards to traffic flow and road usage. Generally there are numerous ways by which road collisions can be investigating leaving the potential future research endless. However, let us just concentrate on the findings of this thesis and where possible directions might be considered. With regard to the typology it would be useful, for example, to disaggregate the density surface to only pedestrian or cyclist collisions in order to determine a more detailed portrayal and typology for traditional vulnerable road user groups only.

Current research being conducted at the Centre for Transport Studies at University College London is investigating the prevalence of child involvement in collisions and deprivation in Greater Manchester. The analysis focuses on deprivation and road collision involvement particularly in children. It is funded by the Neighbourhood Road Safety Initiative, a government initiative to tackle the significantly higher incidence of road traffic injuries in disadvantaged communities. This shows a positive step towards understanding the dynamics of road collisions at a neighbourhood level. Unfortunately the study does not include London in the areas of focus. However this highlights the necessity for temporal analysis, locational hotspots change over time (regression-to-the-mean) and this requires constant monitoring, therefore an up-to-date rolling analysis of the collisions at yearly intervals would be advantageous for any robust initiatives to be implemented. Disaggregation of the collision dataset can be a useful tool in order to create a more localised understanding of the collisions relating to just one variable. For example, I have already mentioned disaggregation by vulnerable road user; another possibility would be by severity and prioritise the fatal and serious collisions into a typology of occurrences.

A key drawback for road collision analysis is data. For road safety analysis to continue to become more comprehensive and successful there needs to be data partnership between interested parties. For example the availability of minor collision data from insurance companies would make road collision analysis more robust and aim to determine collision hotspots more effectively. This would have the added information of who was at fault and where they live. This could provide valuable supplementary information to determining better road safety initiatives. Coupled with this, the make and model and age of the vehicles involved would provide a snapshot of the types of vehicles more likely to be involved in collisions. Although this would potentially be controversial, it would provide valuable information for further analysis.

8.4 Towards a better understanding of road collisions?

Whilst there are vehicles, pedestrians and other road users on the road there will continue to be road traffic collisions. This is not a problem that can be eradicated but instead understood better and managed. This thesis suggests understanding road traffic collisions as a public health problem which is the current paradigm in which road safety is being researched. Let us consider obesity as a public health concern. The majority of the population are aware of the lifestyle choices and which choices pose a risk for the likelihood of falling into this category: we need to eat well, take exercise, and avoid unhealthy foods. Generally speaking there has been much research into the lifestyle of the population and their likelihood of being obese. If we consider the likelihood of being in road collision akin to the likelihood of being obese, does society understand the risks as well? This Ph.D. has posed the hypotheses that lifestyle does have an impact on involvement in road collisions. Many previous research studies have concentrated on a single aspect of a person's lifestyle such as deprivation, age, place of residence and personality. However there has been no research which has tried to encompass all these factors and determine a better understanding of people's road safety risk as a whole proportion of their lifestyle.

London is an extremely diverse city, with a resident population that competes in the road environment on a daily basis. There are people who cycle, walk, use the car for business, use the car for pleasure, motorbike, visitors etc. This interaction of different interests is bound to cause conflict. The sheer density of London, its roads and population mean that unpacking the patterns and processes that exist can be extremely complex.

People's perception of their own risk in the road environment is generally under estimated. You are more likely to die in a road collision than in an airplane crash yet somehow this does not affect how we perceive our interaction with the road environment. This is partly because we interact on a daily basis, but the difference is, that different people in society have different levels of risk. For example young ethnic males driving in old sports cars are likely to be in a road collision. The use of geodemographics to understand and gain insight into crime, health and education has motivated this use in a road safety context. It is not the intention that it answers all the questions of lifestyle and road collision involvement but it offers a clearer picture of the types of people likely to be involved and most importantly, where they are likely to be injured. In the past, the large proportion of research in road safety concentrated on, quite rightly the location of the collision itself. However, this thesis has aimed to show that by only looking at the collision location, there is a lot of information that is being left out of the equation. Although it is impossible to incorporate all the facets relating to collision involvement there needs to be an understanding of the many spatial locations which are involved, particularly where the casualties and drivers reside.

The major difference between road collisions and say, obesity is that there is a large proportion of chaos and uncertainty involved in the likelihood of being involved. Risks of being in a road collision are inherently subjective and incorporate many different players, including the road user, any other road user, the road environment, time, mood, type of car. It would be impossible to accurately capture a person's risk of being in road collision because of the difficulty of unpacking all these competing factors. There are so many facts banded around in the media proclaiming the high numbers of road collision deaths in relation to say something like terrorism. A study at the University of Otago determined that the body count from road collisions in developed economies is 390 times higher than the death toll in these counties from international terrorism. To put this into context, in 2001 as many people died every 26 days on America's roads as died in the terrorist attacks of 9/11.

The study of road safety crosses many different domains, including, engineering, psychology, and medicine but it is really in recent years as they study been researched from a geographical perspective. Now, having this geographical element, there needs to be increased incorporation with other wider disciplines in order to begin a more multifaceted approach to road safety. As the saturation of road safety engineering measures in the built environment come into effect, and as the number of vehicles on the roads increases there is a need to take a different approach to this public health issue. Road safety is largely political. Different countries have different ways of managing and dealing with road collisions, and the research shows that there is no right or wrong way.

The study of road collisions is a problematic discipline, especially when trying to understand it from a larger perspective. The longer one spends in an effort to understand them, the more questions are asked. When analysed in an environment where road collision are measured against say, one or two variables the conclusions are straightforward, however this does not replicate real life. Taken in its real environment, the study of road collisions proves almost overwhelming and almost impossible.

APPENDIX

Symbols of Success
Global Connections
Cultural Leadership
Corporate Chieftains
Golden Empty Nesters
Provincial Privilege
High Technologists
Semi-Rural Seclusion
Happy Families
Just Moving In
Fledgling Nurseries
Upscale New Owners
Families Making Good
Middle Rung Families
Burdened Optimists
In Military Quarters
Suburban Comfort
Close to Retirement
Conservative Values
Small Time Business
Sprawling Subtopia
Original Suburbs
Asian Enterprise
Ties of Community
Respectable Rows
Affluent Blue Collar
Industrial Grit
Coronation Street
Town Centre Refuge
South Asian Industry
Settled Minorities
Urban Intelligence
Counter Cultural Mix
City Adventurers
New Urban Colonists
Caring Professionals
Dinky Developments
Town Gown Transition
University Challenge
Welfare Borderline
Bedsit Beneficiaries
Metro Multiculture
Upper Floor Families
Tower Block Living
Dignified Dependency
Sharing a Staircase
Municipal Dependency
Families on Benefits
Low Horizons
Ex-Industrial Legacy
Blue Collar Enterprise
Rustbelt Resilience
Older Right to Buy
White Van Culture
New Town Materialism
Twilight Subsistence
Old People in Flats
Low Income Elderly
Cared for Pensioners
Grey Perspectives
Sepia Memories
Childfree Serenity
High Spending Elders
Bungalow Retirement
Small Town Seniors
Tourist Attendants
Rural Isolation
Summer Playgrounds
Greenbelt Guardians
Parochial Villagers
Pastoral Symphony
Upland Hill Farmers

Mosaic UK Groups and Types

Mosaic classifies households in the United Kingdom by allocating them to one of 61 types and 11 groups.

Group	Group Description	% Households	Type	Type Description	% Households
A	Symbols of Success	9.62	A01	Global Connections	0.72
			A02	Cultural Leadership	0.92
			A03	Corporate Chieftains	1.12
			A04	Golden Empty Nesters	1.33
			A05	Provincial Privilege	1.66
			A06	High Technologists	1.82
			A07	Semi-Rural Seclusion	2.04
B	Happy Families	10.76	B08	Just Moving In	0.91
			B09	Fledgling Nurseries	1.16
			B10	Upscale New Owners	1.35
			B11	Families Making Good	2.32
			B12	Middle Rung Families	2.86
			B13	Burdened Optimists	1.96
			B14	In Military Quarters	0.17
C	Suburban Comfort	15.10	C15	Close to Retirement	2.81
			C16	Conservative Values	2.84
			C17	Small Time Business	2.93
			C18	Sprawling Subtopia	3.06
			C19	Original Suburbs	2.41
			C20	Asian Enterprise	1.02
D	Ties of Community	16.04	D21	Respectable Rows	2.65
			D22	Affluent Blue Collar	3.12
			D23	Industrial Grit	3.82
			D24	Coronation Street	2.81
			D25	Town Centre Refuge	1.13
			D26	South Asian Industry	0.88
			D27	Settled Minorities	1.62
E	Urban Intelligence	7.19	E28	Counter Cultural Mix	1.36
			E29	City Adventurers	1.27
			E30	New Urban Colonists	1.36
			E31	Caring Professionals	1.06
			E32	Dinky Developments	1.10
			E33	Town Gown Transition	0.76
			E34	University Challenge	0.26
F	Welfare Borderline	6.43	F35	Bedsit Beneficiaries	0.71
			F36	Metro Multiculture	1.67
			F37	Upper Floor Families	1.72
			F38	Tower Block Living	0.40
			F39	Dignified Dependency	1.34
			F40	Sharing a Staircase	0.50
G	Municipal Dependency	6.71	G41	Families on Benefits	1.21
			G42	Low Horizons	2.64
			G43	Ex-Industrial Legacy	2.86
H	Blue Collar Enterprise	11.01	H44	Rustbelt Resilience	3.00
			H45	Older Right to Buy	2.67
			H46	White Van Culture	3.17
			H47	New Town Materialism	2.17
I	Twilight Subsistence	3.88	I48	Old People in Flats	0.83
			I49	Low Income Elderly	1.63
			I50	Cared for Pensioners	1.43
J	Gray Perspectives	7.88	J51	Sepia Memories	0.75
			J52	Childfree Serenity	1.34
			J53	High Spending Elders	1.53
			J54	Bungalow Retirement	1.26
			J55	Small Town Seniors	2.71
			J56	Tourist Attendants	0.30
K	Rural Isolation	5.39	K57	Summer Playgrounds	0.29
			K58	Greenbelt Guardians	1.74
			K59	Parochial Villagers	1.64
			K60	Pastoral Symphony	1.31
			K61	Upland Hill Farmers	0.41

mosaic
United Kingdom

Figure 1*: Mosaic UK Types and Groups*

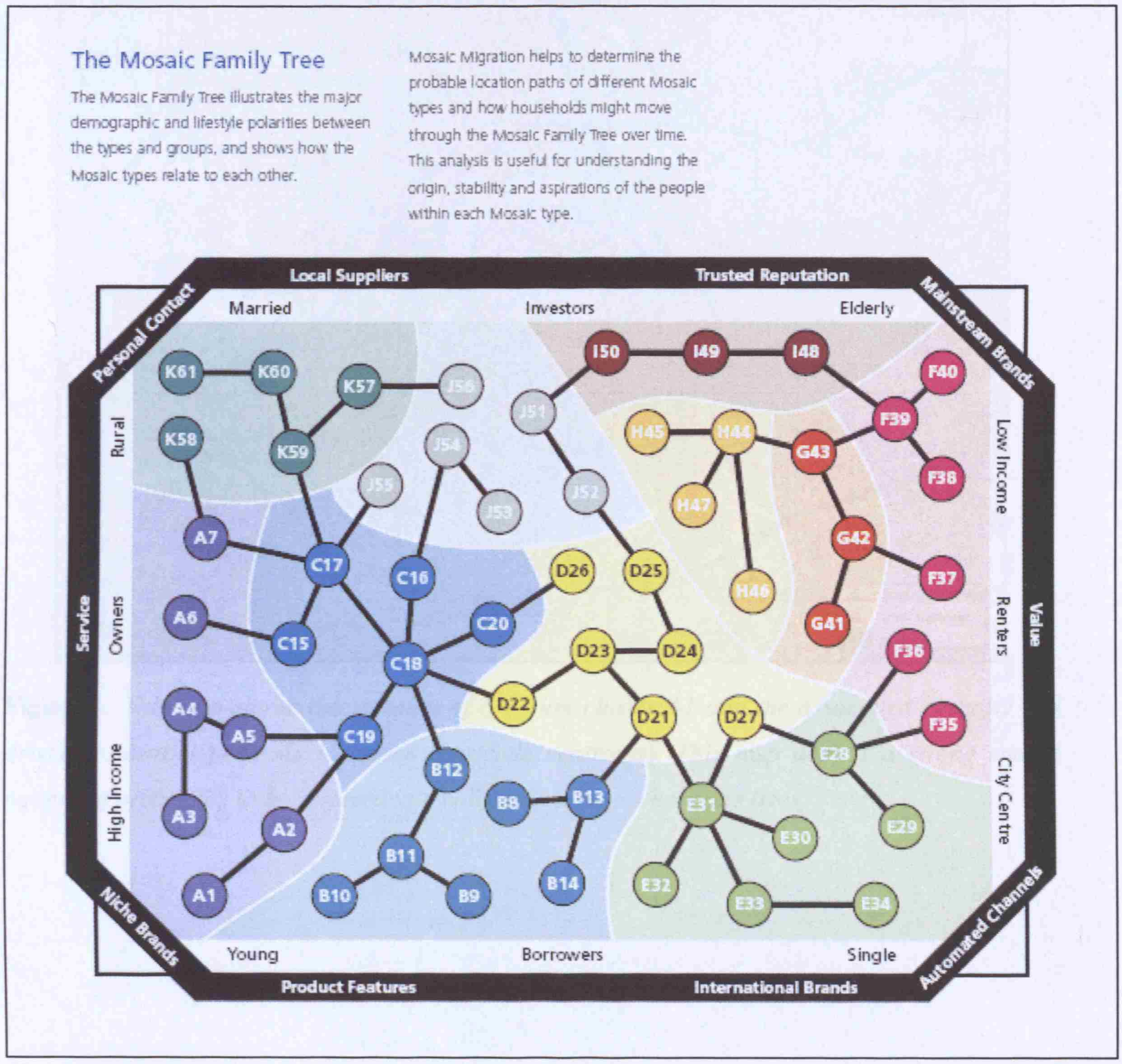


Figure 2: *The Mosaic Family Tree*

Figure 3: *This map shows the location of collision cluster A1 and the associated casualty and driver residential postcode locations (postcode centroids). This map depicts a strong spatial pattern of propensity to be involved in a collision close to where one lives*

.

Figure 4: *This map shows the location of collision cluster A5 and the associated casualty and driver residential postcode locations (postcode centroids). This map depicts a strong spatial pattern of propensity to be involved in a collision close to where one lives.*

Figure 5: *This map shows the location of collision cluster D14 and the associated casualty and driver residential postcode locations (postcode centroids). This map depicts a strong spatial pattern of propensity to be involved in a collision close to where one lives.*

Mosaic TYPE	Collision Clusters more likely to be involved in (driver)	Collision Clusters more likely to be involved in (driver)
A01 Global Connections	D14, D13, A8	C10, B9
A02 Cultural Leadership	A8, D13, D14	-
A03 Corporate Chieftains	A8, A7	-
A04 Golden Empty Nesters	-	A7
A05 Provincial Privilege	C10, A8	-
A06 High Technologists	A4	E15, D14, A4
A07 Semi Rural Seclusion	A6, A4	B9, D11
B08 Just Moving In	-	-
B09 Fledging Nurseries	A2, A4, A5	D14, D12, D13
B10 Upscale New Owners	A6, A5, A4	A6, A4, D12
B11 Families Making Good	A2	B9, D13
B12 Middle Rung Families	-	E15, D13, D14
B13 Burdened Optimists	A2, A5	D14, D12, D13
B14 Military Quarters	-	D13
C15 Close to Retirement	-	D13
C16 Conservative Values	-	D13
C17 Small Time Business	A2, A6	D13, D14, D12
C18 Sprawling Subtopia	C10, A8	E15
C19 Original Suburbs	A8, C10	D11
C20 Asian Enterprise	C10, A8	A1, D14
D21 Respectable Rows	A8, C10	E15
D22 Affluent Blue Collar	A6	D12, D14, A7
D23 Industrial Grit	A5	D14, D11
D24 Coronation Street	-	A8
D25 Town Centre Refuge	A2	A2, A7
D26 South Asian Industry	A1	A1, A7
D27 Settled Minorities	A8, C10, A1	A3, D13, D11
E28 Counter Cultural Mix	D13, D14	B9, C10
E29 City Adventurers	D13, D14, D12	A6, C10
E30 New Urban Colonists	A3, A8, D13	A3, B9
E31 Caring Professionals	A5	A6
E32 Dinky Developments	A2	E15, D13, D14
E33 Town Gown Transition	-	C10
E34 University Challenge	-	-

F35 Bedsit Beneficiaries	A6	D14, A8
F36 Metro Multiculture	D13, A8. C10	A1, A3
F37 Upper Floor Families	A4	D11, E15
F38 Tower Block Living	A5	A1
F39 Dignified Dependency	A5	A7, C10
F40 Sharing a Staircase	A5	A7, A5
G41 Families on Benefits	A3	A3, E15
G42 Low Horizons	A2, A5	D13
G43 Ex-Industrial Legacy	A2	D14, B9
H44 Rustbelt Resilience	A5	D13
H45 Older Right to Buy	A6, A4	D14, A4, A6
H46 White Van Culture	A8, C10, D11	D13, D14
H47 New Town Materialism	A4, A2	D14, A4
I48 Old People in Flats	-	A7
I49 Low Income Elderly	-	E15
I50 Cared For Pensioners	-	D13
J51 Sepia Memories	A8	D14
J52 Childfree Serenity	A8, D13, C10	E15
J53 High Spending Elders	A6	D12
J54 Bungalow Retirement	-	D14
J55 Small Town Seniors	A5	E15
J56 Tourist Attendants	-	-
K57 Summer Playgrounds	-	-
K58 Greenbelt Guardians	A5, A4, A2	D13, A2
K59 Parochial Villagers	A6	A7, D12
K60 Pastoral Symphony	-	-
K61 Upland Hill Farmers	-	-

Table 9.1: *Table to highlight which Mosaic Types are more likely to be in certain collision Clusters.*

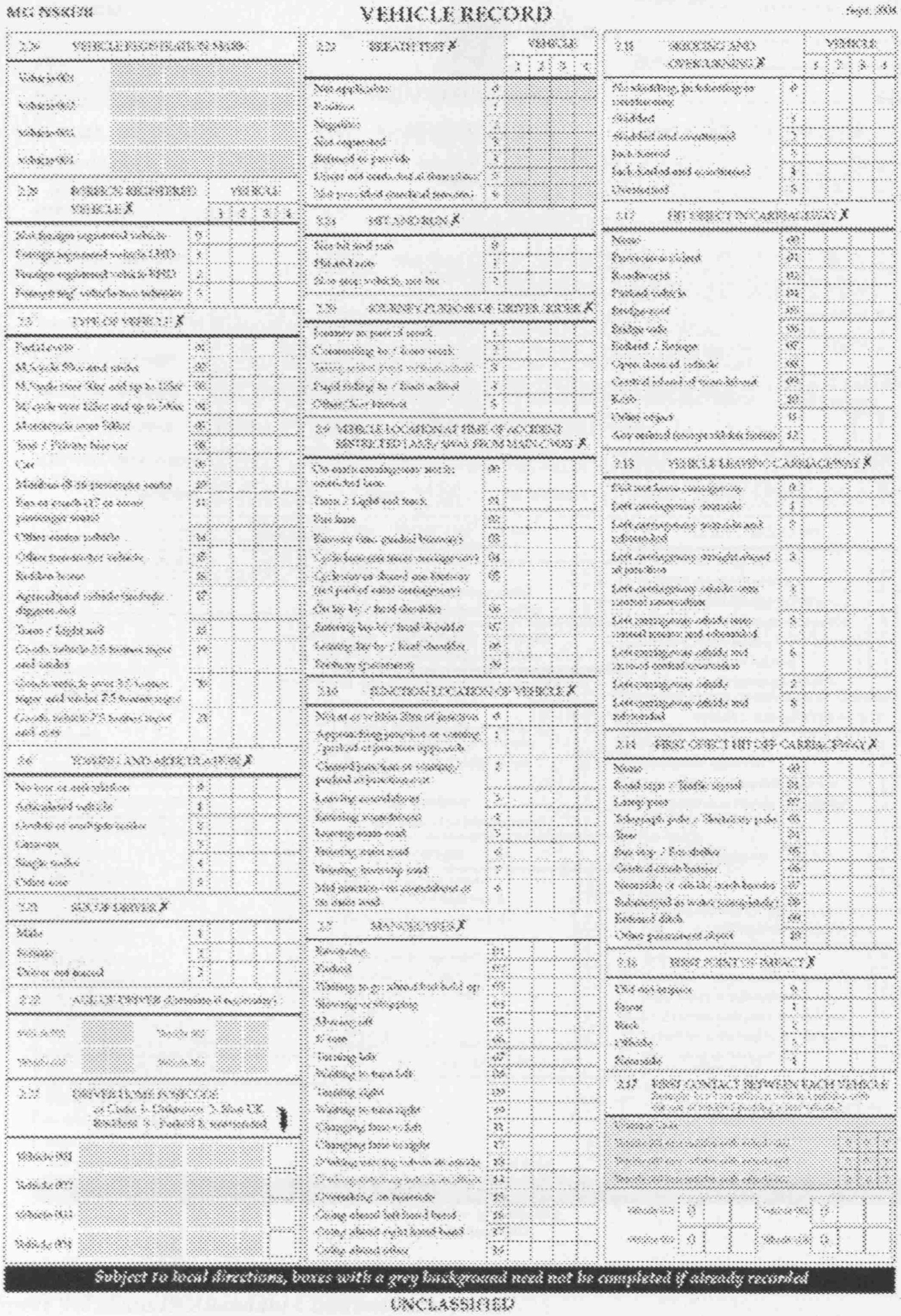
Mosaic TYPE	Collision Group(s) most likely to be involved in (driver)	Collision Group(s) most likely to be involved in (casualty)
A01 Global Connections	C	C
A02 Cultural Leadership	C	-
A03 Corporate Chieftains	-	-
A04 Golden Empty Nesters	-	-
A05 Provincial Privilege	C	-
A06 High Technologists	D	E
A07 Semi Rural Seclusion	D	B
B08 Just Moving In	-	-
B09 Fledging Nurseries	D	D
B10 Upscale New Owners	A	D
B11 Families Making Good	D	B
B12 Middle Rung Families	D	E
B13 Burdened Optimists	D	D
B14 Military Quarters	-	-
C15 Close to Retirement	D	D
C16 Conservative Values	-	-
C17 Small Time Business	D	D
C18 Sprawling Subtopia	C	E
C19 Original Suburbs	C	-
C20 Asian Enterprise	C	E
D21 Respectable Rows	C	E
D22 Affluent Blue Collar	D	D
D23 Industrial Grit	D	D
D24 Coronation Street	D	-
D25 Town Centre Refuge	D	D
D26 South Asian Industry	D	A
D27 Settled Minorities	C	D
E28 Counter Cultural Mix	C	B
E29 City Adventurers	A/C	C
E30 New Urban Colonists	C	B
E31 Caring Professionals	A	B
E32 Dinky Developments	D	E
E33 Town Gown Transition	-	C
E34 University Challenge	-	-
F35 Bedsit Beneficiaries	D	A

F36 Metro Multiculture	C	C
F37 Upper Floor Families	D	E
F38 Tower Block Living	-	B
F39 Dignified Dependency	-	C
F40 Sharing a Staircase	-	-
G41 Families on Benefits	D	E
G42 Low Horizons	A	D
G43 Ex-Industrial Legacy	D	B
H44 Rustbelt Resilience	D	D
H45 Older Right to Buy	D	D
H46 White Van Culture	C	D
H47 New Town Materialism	D	D
I48 Old People in Flats	-	-
I49 Low Income Elderly	D	E
I50 Cared For Pensioners	-	-
J51 Sepia Memories	-	-
J52 Childfree Serenity	C	E
J53 High Spending Elders	-	D
J54 Bungalow Retirement	-	D
J55 Small Town Seniors	-	E
J56 Tourist Attendants	-	-
K57 Summer Playgrounds	-	-
K58 Greenbelt Guardians	D	A
K59 Parochial Villagers	D	C
K60 Pastoral Symphony	-	-
K61 Upland Hill Farmers	-	-

Table 9.2: *Table represents the collision Groups which each Mosaic Type is more likely to involved in*

COLLISION CLUSTER	Mosaic Type (Driver)	Mosaic Type (Casualty)
A1	D26, D27, F36	D26, F36, C20
A2	E32, D27	D27, E28, E32
A3	E30, F36, D27	E30, D27, F36
A4	H47, D27, F36	F36, D27
A5	F36, E28	E28, F36
A6	D26, F36, E28	E29, F36, E28
A7	C19, D27, C20	D26, F36, E28
A8	C19, D27, H46	F36, E28
B9	N/A	E28, E30
C10	D27, C19, C20	E28, E29, F36
D11	H46, A01	F36, D27
D12	F36, E28, E29	H46, C15, B12
D13	E28, A01, F36	H46, B12, C20
D14	A01, E28, E29	H47, H46, C20
D15	N/A	C18, D21, B12

Table 9.3: *This table shows the collision clusters and associated Mosaic Types (for driver and casualty) that are more likely to be involved in a collision. The first Mosaic Type against each cluster is the top scoring index score (highly over represented). More detailed information about this can be found in Chapter 7.*



VEHICLE RECORD

Subject to local directions, boxes with a grey background need not be completed if already recorded

UNCLASSIFIED

Figure 9.6: *Stats19 Vehicle form*

MG NSRFA

ACCIDENT STATISTICS

1.2 ACCIDENT REFERENCE

*FATAL / SERIOUS / SLIGHT

1.3 TIME DAY* Su M T W Th F S 1.7 DATE 2 0

1st Road Class & No. or (Unclassified - UC) (Not Known - NK) 1st Road Name

Outside House No. or Name or Marker Post No. at junction with / or metres N S E W * of

2nd Road Class & No. or (Unclassified - UC) (Not Known - NK) 2nd Road Name

Town Sector /Beat No.

County or Borough

Parish No. or Name 1.10 Local Auth No. (if known)

1.11 Grid Reference E → N ↑

REPORTING OFFICER Name Number

BCU/Stn 1.1 Force Tel Number

1.5 Number of vehicles	
1.6 Number of casualties	

1.13 ROAD TYPE

Roundabout	1
One way street	2
Dual carriageway	3
Single carriageway	6
Slip road	7
Unknown	9

1.15 Speed Limit (Permanent)

1.16 JUNCTION DETAIL

Not at or within 20 metres of junction	00
Roundabout	01
Mini roundabout	02
T or staggered junction	03
Slip road	05
Crossroads	06
Multiple junction	07
Using private drive or entrance	08
Other junction	09

JUNCTION ACCIDENTS ONLY

1.17 JUNCTION CONTROL

Authorised person	1
Automatic traffic signal	2
Stop sign	3
Give way or uncontrolled	4

1.20a PEDESTRIAN CROSSING - HUMAN CONTROL

None within 50 metres	0
Control by school crossing patrol	1
Control by other authorised person	2

1.20b PEDESTRIAN CROSSING - PHYSICAL FACILITIES

No physical crossing facility within 50m	0
Zebra crossing	1
Pelican, puffin, toucan or similar non-junction pedestrian light crossing	4
Pedestrian phase at traffic signal junction	5
Footbridge or subway	7
Central refuge — no other controls	8

1.22 WEATHER

Fine without high winds	1
Raining without high winds	2
Snowing without high winds	3
Fine with high winds	4
Raining with high winds	5
Snowing with high winds	6
Fog or mist — if hazard	7
Other	8
Unknown	9

1.23 ROAD SURFACE CONDITION

Dry	1
Wet / Damp	2
Snow	3
Frost / Ice	4
Flood (surface water over 3cm deep)	5

1.21 LIGHT CONDITIONS

Daylight: street lights present	1
Daylight: no street lighting	2
Daylight: street lighting unknown	3
Darkness: street lights present and lit	4
Darkness: street lights present but unlit	5
Darkness: no street lighting	6
Darkness: street lighting unknown	7

1.24 SPECIAL CONDITIONS AT SITE

None	0
Auto traffic signal out	1
Auto traffic signal partially defective	2
Permanent road signing or marking defective or obscured	3
Roadworks	4
Road surface defective	5
Oil or Diesel	6
Mud	7

1.25 CARRIAGEWAY HAZARDS

None	0
Dislodged vehicle load in carriageway	1
Other object in carriageway	2
Involvement with previous accident	3
Pedestrian in carriageway - not injured	6
Any animal in carriageway (except ridden horse)	7

1.26 Did a police officer attend the scene and obtain the details for this report?

Yes	1
No	2

Subject to local directions, boxes with a grey background need not be completed if already recorded

* Circle as appropriate

UNCLASSIFIED

Figure 9.7: *Stats19 Attendant Circumstances form*

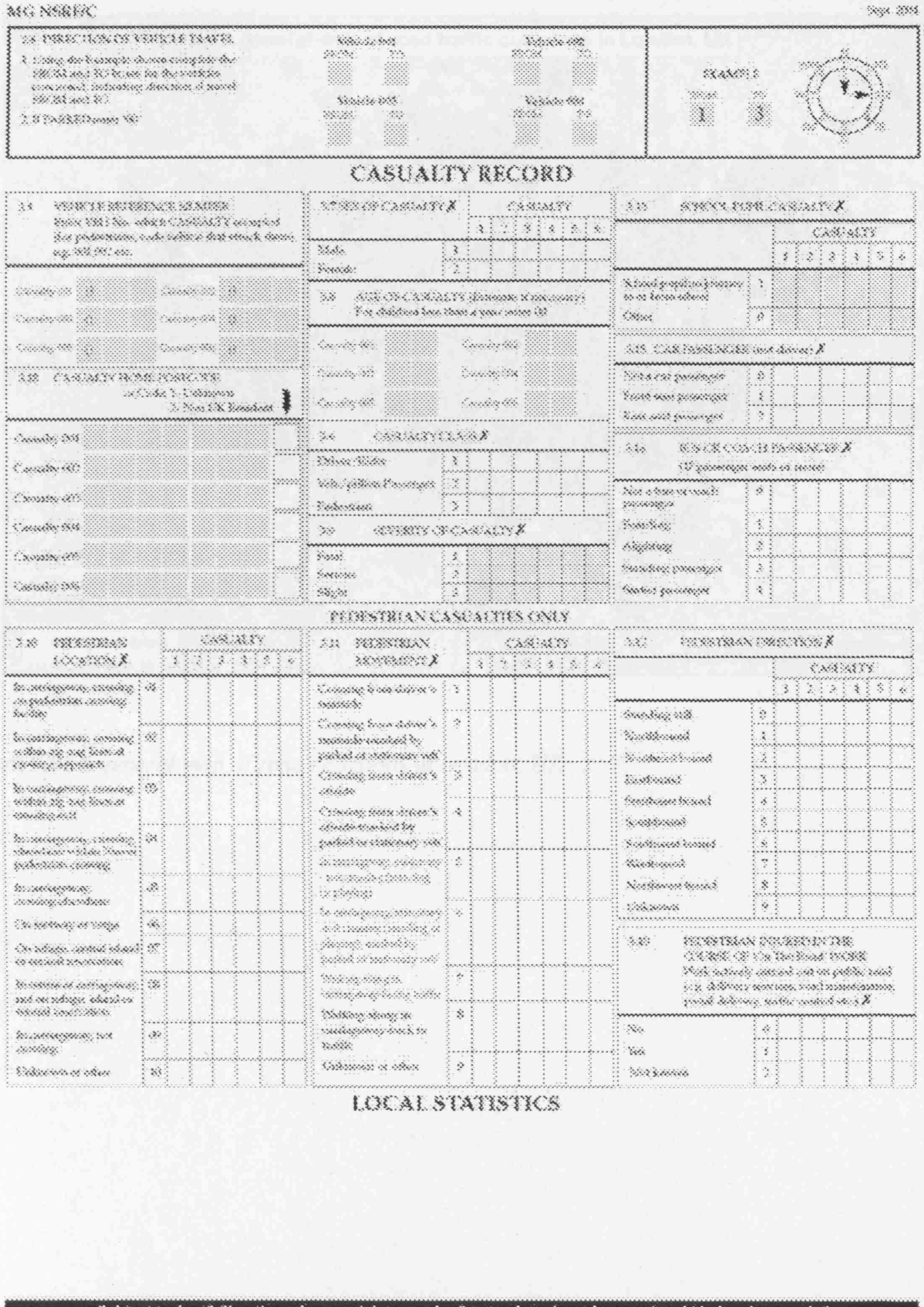


Figure 9.8: Stats19 Casualty form

Figure 9.9: *A general map of road collisions in London, UK*

Figure 9.10: *A kernel density estimation map of road traffic collisions in London, UK*

Variable	Comments	Used in Clustering Process
Accident key	Not very relevant for study	N
No. of vehicles	Important, determines scale of collision	Y
No. of casualties	Important, determines scale and magnitude of collision	Y
Record Type	Not relevant for the study	N
Crash Reference	Not relevant for this study	N
Police Force	Not analysing the police force areas, as looking at general spatial location	N
Severity of accident	Good general indicator for the hotspot	Y
Date	Not hugely relevant as not doing long term temporal analysis	Y
Time	Found inconsistencies in terms of five minute slots – had to create time bands	Y
Day of week	Useful for determine weekday/weekend pattern of involvement	Y
Local Authority	Not relevant as not studying local authority criteria	N
Easting	Spatial reference to nearest 10 metres – deemed accurate	Y
Northing	Spatial reference to nearest 10 metres – deemed accurate	Y
First road class	Not used – road map used instead	N
First road number	Not used – road map used instead	N
Road type	Could deem this information from road map in a spatial context	N
Speed limit	Not seemed relevant in a city where predominant speed limits are low and un variable	N
Junction detail	Could deem this information from road map	N
Junction control	Could deem this information from road map	N
Second road class	Could deem this information from road map	N
Second road number	Could deem this information from road map	N
Pedestrian crossing – human control	Disregarded in favour of specific location of pedestrian crossing for the clustering	N
Pedestrian crossing – physical control	Disregarded in favour of specific location of pedestrian crossing for the clustering	N
Light conditions	Too many options to include and weather was deemed more important	N
Weather	Good, although majority of collisions occurred in fine weather or weather was 'unknown'.	Y
Road surface conditions	Deemed to be similar to weather and a large proportion was either dry or unknown	N
Special conditions	Not to be investigated, included items such as road works	N
Carriageway hazards	Deemed not relevant, includes animals or vehicles in roadway	N
Place reported	Not deemed relevant	N
DTLR Special	Not deemed relevant or sufficient information	N
Contributory Factors	Deemed unreliable by police – not used	N
Police collision	Not relevant for study	N
Location description	No 'key' words to be able to search. Too ambiguous	N
Crash description	No 'key' words to be able to search. Too ambiguous	N

Figure 9.11: *Stats19 data and 'fitness for use' summary – Attendant Circumstances*

Variable	Comments	Used in clustering process
Accident Key	Not deemed relevant for this study	N
Crash reference	Not deemed relevant for this study	N
Casualty ref. number	Not deemed relevant for this study	N
Casualty record	Not deemed relevant for this study	N
Vehicle reference	Not deemed relevant for this study	N
Class of casualty	Extremely important. Determines whether the casualty is pedestrian, cyclist, driver etc.	Y
Sex of casualty	Important, although not included at this stage. Wanted to gain a 'physical' indication of hotspot patterns	N
Age of casualty	Important, although not included at this stage. Wanted to gain a 'physical' indication of hotspot patterns	N
Severity of casualty	Important, but wanted to gauge an indication of the collisions as a whole rather than individuals at the hotspot detection level	N
Pedestrian location	Related to individuals rather than the collision as a whole	N
Pedestrian movement	Related to individuals rather than the collision as a whole	N
Pedestrian direction	Related to individuals rather than the collision as a whole	N
School pupil casualty	Related to individuals rather than the collision as a whole	N
Car passenger	Related to individuals rather than the collision as a whole	N
PSV Passenger	Related to individuals rather than the collision as a whole	N
DTLR special projects	Related to individuals rather than the collision as a whole	N
Casualty postcode	An extremely important spatial reference for casualties involved, used for the initial geodemographic analyses rather than the hotspot clustering	N

Figure 9.12: *Stats19 data and 'fitness for use' summary – Casualty Details*

Variable	Comments	Used for clustering process
Accident key	Not deemed relevant for this study	N
Crash reference	Not deemed relevant for this study	N
Vehicle reference	Not deemed relevant for this study	N
Record type	Not deemed relevant for this study	N
Vehicle type	Important, but wanted to maintain the basic distinction between cyclists, pedestrians and car casualties	N
Towing	Not deemed relevant for this study	N
Manoeuvres	Important, but linked to the individual vehicle rather than the collision as a whole entity	N
Vehicle movement from	Not deemed relevant for this study	N
Vehicle movement to	Not deemed relevant for this study	N
Vehicle location road	Not deemed relevant for this study	N
Vehicle location away	Not deemed relevant for this study	N
Junction location	Not deemed relevant for this study	N
Skid	Not deemed relevant for this study	N
Hit object in carriageway	Would include animals or other vehicles	N
Vehicle leaving carriageway	These detailed circumstances are focused to the very small scale engineering problems of hotspots	N
Hit object off carriageway	These detailed circumstances are focused to the very small scale engineering problems of hotspots	N
First point of impact	This links to the individual circumstances rather the collision as a whole	N
Other vehicles hit	Deemed not relevant, preferred to use number of vehicles involved	N
Part damaged 1	Deemed not relevant for this study	N
Parts damaged 2	Deemed not relevant for this study	N
Parts damaged 3	Deemed not relevant for this study	N
Sex of driver	These were not deemed a relevant part of the physical analysis of hotspot clusters – it was added later	N
Age of driver	These were not deemed a relevant part of the physical analysis of hotspot clusters – it was added later	N
Breath test	An interesting variable, but very few drivers where tested and even fewer were positive	N
Hit and run	An interesting variable but due to the lack of data collection, it was not used	N
DTLR Special circumstances	Not deemed relevant	N
Driver postcode	An important spatial component. Used for the preliminary geodemographic analyses rather than the clustering	N

Figure 9.13: Stats19 data and 'fitness for use' summary – Vehicle details

BIBLIOGRAPHY

Adams, J. (1995) Risk, Routledge, London, UK

Adams, J. (1999) Risky Business, The Adam Smith Institute, London

AA Trust (2003) The facts about road accidents and children, Available on the World Wide Web, URL http://www.aamotoringtrust.com (Accessed 14th June 2003)

Abdalla, I.M & Raeside, R. & Barker, D. and McGuigan, D. (1997) An investigation into the relationships between areal social characteristics and road accident casualties, Accident Analysis and Prevention Vol. 29 No. 5 pp 583–593

Abdalla, I. M. (1999) Linking road accident statistics to census data, PhD Thesis, Napier University

Abdel-Aty, M.A. and A.E. Radwan (2000) Modelling traffic accident occurrence and involvement Accident Analysis and Prevention, Vol. 32 No.5 pp. 633-642

ACPO (2003) Modern Road Policing: A Manifesto for the future, Available on the World Wide Web, URL http://www.acpo.police.uk (Accessed 28th July 2004)

Agent, K. and Deen, R. (1975) Relationships between roadway geometrics and accidents, Transportation Research Board, Issue No. 541

Akerstedt, T. and Kecklund, G. (2001) Age, gender and early morning highway accidents, Journal of Sleep Research, Vol. 10 pp. 105-110

Ameen, J and Naji, J. (2001) Causal models for road accident fatalities in Yemen, Accident Analysis and .Prevention Vol. 33 No. 4 pp. 547-561

Anas, A., Arnott, R. and Small, K. (1998) Urban spatial structure, Journal of Economic Literature, Vol. 36 pp. 1426–1464.

Anderson (2000) Spatial analysis of road accidents in Warwickshire, Undergraduate Thesis, University College London

Andrey, J. and Olley, R., (1990) Relationships between weather and road safety, past and future directions. Climatology Bulletin Vol. 24 No.3 pp. 123–137.

Armheim, C. (1995) Searching for the elusive aggregation effect: Evidence from statistical simulations, Environment and Planning A, Vol. 27 No. 1 pp. 105

Ashby, D. and Longley, P. (2005) Geocomputation, Geodemographics and Resource Allocation for Local Policing, Transactions in GIS, Vol. 9 No. 1 pp. 53-72

Atkinson, P. and Tate, N. (2000) Spatial Scale Problems and Geostatistical Solutions: A Review, Professional Geographer, Vol. 52 No. 4 pp. 607-623

Austin, K. (1995) The identification of mistakes in road accident records: Part 1, Location Variables, Accident Analysis and Prevention, Vol. 27 No. 2 pp. 261-276

Austin, K. (1995) The identification of mistakes in road accident records: Part 2, Casualty Variables, Accident Analysis and Prevention, Vol. 27 No. 2 pp. 277-282

Baker *et al* (2003) Older women drivers: Fatal crashes in good conditions, Journal of Safety Research Vol. 34 No. 4 pp. 399-405

Bailey, T.C. and Gatrell, A. C. (1995) Interactive Spatial Data Analysis. John Wiley and Sons, New York, NY

Batey, P. and Brown, P. (1995) From human ecology to customer targeting: the evolution of geodemographics, in Longley, P. and Clarke, G.P., (eds.) GIS for Business and Service Planning, GeoInformation, Cambridge

Batty, M. & de Almeida & Camera, G. & Cerqueira, C.G. & Monteiro, A. M.W & Pennachin, C.P. & Soares-Filho, B.S. (2003) Stochastic Cellular Automata Modelling of Urban Land Use Dynamics: Empirical Development and Estimation, Computers, Environment and Urban Systems, Vol. 27 pp.481-509

Batty, M. and Longley, P. (1995) Fractal Cities: A Geometry of Form and Function Academic, San Diego, 1994

Beck, U. (1992) Risk Society: Towards a New Modernity, Sage, London

Bédard *et al* (2002) The independent contribution of driver, crash and vehicle characteristics to driver fatalities, Accident Analysis and Prevention, Vol. 34 No. 6 pp. 717-727

Birkin, M. & Clarke, G. P. (1998) GIS, geodemographics and spatial modelling: an example within the UK financial service industry, Journal of Housing Research Vol. 9 No.1 pp. 87-111

Black, W. and Thomas, I. (1998) Accidents on Belgium's motorways: a network autocorrelation analysis, Accident Analysis and Prevention, Vol. 6. No. 1 pp. 23-31

Blatt, J. and Furman, S. (1998) Residence location of drivers involved in fatal crashes, Accident Analysis and Prevention, Vol. 30 No. 6 pp. 705-711

Blower, D. and Campbell, K. and Green, P. (1993) Accident rates for heavy truck-tractors in Michigan Accident Analysis and Prevention Vol. 25 No.3 pp. 307-321.

Borruso, G. (2003) Network density and the delimitation of urban areas, Transactions in GIS, Vol. 7 No. 2 pp. 177-191

Bort, H. (2004) Factorial structure of recklessness: To what extent are older drivers different?, Journal of Safety Research, Vol. 35 No. 3 pp. 329-335

Brijs, T. and Swinnen, G. and Vanhoof, K. and Wets, G. (2004): Building an association rules framework to improve product assortment decisions Data Mining and Knowledge Discovery, Vol. 8 No.1 pp.7-23.

Brodsky, H. and Hakkert, A.S. (1988) Risk of a road accident in rainy weather, Crash Analysis Prevention Vol. 20 No.2 pp. 161–176.

Broughton, J. and Hazelton, M. and Stone, M. (1999) Influence of light level on the incidence of road casualties and predicted effect of changing 'summertime', Journal of Royal Statistical Society A, Vol. 162 No. 2 pp. 137-175

Brown, P.J.B. (1991) Exploring geodemographics In I. Masser and M. Blakemore (eds), Geographic Information Management: Methodology and Applications, pp. 221-258, Longman, London

Brown, PJB, Hirschfield, AFG and Batey, PWJ (2000) Adding value to census data: public sector applications of Super Profiles geodemographic typology, *Journal of Cities and Regions*, X, 19-32

Buttenfield, B. and Weibel, S. (2001) Data quality and Uncertainty, Conference Proceedings for the International Conference on Geographic Information Science

Carlstein, T. and Parkes, D. and Thrift N. (1978) (eds) Spacing Time and Timing Space, Arnold, London

Carr, D. and Dyte, D. and Cameron, M. (1995) Analysis of high risk and high severity groups among motorcyclists, Monash University Accident Research Centre – Report 77

Ceder, A. and Livneh, M. (1982) Relationships between road accidents and hourly traffic flow Accident Analysis and Prevention Vol.14 No.1 pp. 19-34.

Cenek, P. and Davies, R. (2000) Crash risk relationships for improved safety management of roads, Avaliable on the World Wide Web, URL http://www.robertnz.net (Accessed 26th November 2004)

Chainey, S.P., Reid S. and Stuart, N. (2002) When is a hotspot a hotspot? A procedure for creating statistically robust hotspot maps of crime. D. Kidner, G. Higgs, and S. White (Eds.) 'Innovations in GIS 9: Socio-economic applications of geographic information science'. London: Taylor and Francis

Chainey, S. and Ratcliffe, J. (2005) GIS and Crime Mapping, John Wiley and Sons, UK

Chapin, F.Stuart Jr. (1974) Human Activity Patterns in the City: Things People Do in Time and in Space. N.Y.: Wiley

Chichester, B.M & Gregan, J.A. & Anderson, D. P. & Kerr, J.M (1998) Associations between road traffic accidents and socio-economic deprivation on Scotland's west coast, Scottish Medical Journal, Vol. 43 No. 5 pp.135-138

Chipman, M. (1995) Risk factors for injury: similarities and differences for traffic crashes and other causes, Accident Analysis and Prevention, Vol. 27 No. 5 pp. 699-706

Chipman, M. and Payne, J. and McDonough, P. (1998) To drive or not to drive: the influence of social factors on the decisions of elderly drivers, Accident Analysis and Prevention, Vol. 30 No. 3 pp. 299-304

Christie, N. (1995) Road accident involvement of children from ethnic minorities, Department for Transport, No 19

Claret, P. and Castillo, J. and Moleon, J. and Cavanillas, A. and Martin, M. and Vargas, R. (2003) Age and sex differences in the risk of causing vehicle collisions in Spain, 1990 to 1999, Accident Analysis and Prevention, Vol. 35 pp. 261-272

Cliff, A. and Ord, J. (1973) Spatial Autocorrelation, Pion, London
Cook, W. and Kazzakov, A. and Persaud, B. (2001) Prioritising highway accident sites: a data enveloping analysis model, Journal of Operational Research Society, Vol. 52 No. 3 pp. 303-309

Cothren, J. (2005) GEOG4863: Quantitative Methods in Geosciences, Centre for Advanced Spatial Technologies, University of Arkansas

Coombes, M. and Raybould, S. (2000) Policy-relevant surfaced data on population distribution and characteristics, Transactions in GIS, Vol. 4 No. 4 pp. 319-342

Cooper, H. and Smaje, C. and Arber, S. (1998) Use of health services by children and young people according to ethnicity and social class: secondary analysis of a national survey British Medical Journal, Vol. 317 pp. 1047-1051

Corfitsen, M. (1999) 'Fatigue' among male night time car drivers: is there a risk taking group?, Safety Science, Vol. 33 No. 1 pp. 47-57

Davies, G. and Patel, D. (2005) The influence of car and driver stereotypes on attributions of vehicle speed, position on the road and culpability in a road accident scenario, Legal and Criminological Psychology, Vol. 10 No. 1 pp. 45-62

Deffenbacher *et al* (2003) Anger, aggression, risky behaviour, and crash related outcomes in three groups of drivers, Behaviour research and Therapy, Vol. 41 No. 3 pp. 333-349

Department of Transport (2000) Road safety and child development research: a summary analysis, Road Safety Research Report No. 5

Department of Transport (2001) Older people: Their transport needs requirements – Main report

Department of Transport (2003) Child road safety in rural areas: literature review and commentary, Road Safety Research Report No. 32

Department of Transport (2005) Child Pedestrian Exposure and Accidents – Further Analyses of Data from a European Comparative Study, Road Safety Research Report No. 56

Department of Transport (2005) Towards the year 2010: monitoring casualties in Greater London, London Accident and Analysis Unit, Issue 5

De Silva, H. (1942) Why we have automobile accidents, John Wiley and Sons, New York

Dissanayake, S. (2003) Young drivers and run-off-the-road crashes, Proceedings of the 2003 Mid-Continent Transportation Research Symposium, Ames, Iowa, Dobson *et al* (1999) Woman drivers' behaviour, socio economic characteristics and accidents, Accident Analysis and Prevention, Vol. 31 No. 4 pp. 525-535

Dobson *et al* (2004) In Australia are people born in other countries at higher risk of road trauma than locally born people?, Accident Analysis and Prevention, Vol. 36 No. 3 pp. 375-381

Donaldson *et al* (2006) Crossing county lines: The impact of crash location and driver's residence on motor vehicle crash fatality, Accident Analysis and Prevention, In press

Dulisse, B. (1997) Older drivers and risk to other road users, Accident Analysis and Prevention, Vol. 29 No. 5 pp. 573-582

Eckhardt, N. and Thomas, I. (2005) Spatial nested scales for road accidents in the periphery of Brussels, IATSS Research, Vol. 29 No. 1 pp. 66-78

Edwards, J. (1996) Weather-related road accidents in England and Wales: a spatial analysis, Journal of Transport Geography, Vol. 4 No. 3 pp. 201-212

Eisenberg, D. (2004) The mixed effects of precipitation on traffic crashes, Accident Analysis and Prevention, Vol. 36 No. 4 pp. 637-647

Elvik, R. (1999) Incomplete accident reporting: Meta-analysis of studies made in 13 countries, Transportation Research Board, No 78 pp. 133-140

Elvik, R. (2003) How would setting policy priorities according to cost–benefit analyses affect the provision of road safety?, Accident Analysis and Prevention, Vol. 35 No. 4 pp. 557-570

Elvik, R. (2004) To what extent can theory account for the findings of road safety evaluation studies?, Accident Analysis and Prevention, Vol. 36 No. 5 pp. 841-849

Elvik, R. (2006) Laws of accident causation, Accident Analysis and Prevention, In press

Flahaut, B (2004) Impact of infrastructure and local environment on road unsafety
Logistic modelling with spatial autocorrelation, Accident Analysis and Prevention, Vol. 36 No. 6 pp. 1055-1066

Flahaut *et al* (2003) The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach, Accident Analysis and Prevention, Vol. 35 No. 6 pp. 991-1004

Flowerdew, R. and Leventhal, B. (1998) Under the microscope, New Perspectives, No. 18 pp. 36-38

Foldvary, L. (1979) Road accident involvement per miles travelled Accident Analysis and Prevention Vol. 11 pp. 75-99.

Folkard, S. (1997) Black times: temporal determinants of transport safety Accident Analysis and Prevention Vol. 29 No. 4 pp. 417-430

Fotheringham, S. and Brunsdon, C. and Charlton, M. (2000) Quantitative Geography: Perspectives on Spatial Data Analysis Thousand Oaks, CA: Sage.

Fountaine, H. and Gourlet, Y. (1997) Fatal pedestrian accidents in France: Typology analysis, Accident Analysis and Prevention, Vol. 29 No. 3 pp. 303-312

Friedman, Michael (1983) Foundations of Time- Space Theories. Princeton Univ. Press

Furedi, F. (1997) Conflict of values and the consciousness of risk, Society for Risk Analysis, Europe 1997 Annual Meeting

Gasser, T. and Kneip, A. and Kohler, W. (1991) Flexible and Fast Method for Automatic Smoothing, J. Amer. Stat. Assoc., Vol. 86 No. 415 pp. 643-652

Geurts, K. and Wets, G. (2003) Blackspot Analysis Methods: Literature Review, Kennis Verkeersonveiligheid, Belgium

Guerts, K. & Thomas, I. & Wets, G. (2005) Understanding the spatial concentrations of road accidents using frequent item sets, Accident Analysis and Prevention, Vol 37 Issue 4 pp. 787-799

Gharaybeh, F.A. (1991) Identification of Accident-Prone Locations in Greater Amman, Transportation Research Record No. 1318 pp. 70-74

Gimotty, P. and T. Chirachavala (1982) Models for the prediction of severe injury 26th Annual Proceedings, American Association for automotive medicine, Ottowa, Ontario, pp. 63-76

Gladwell, M. (2001) Wrong turn: How the fight to make America's roadways safer went off course. New Yorker, 50–61

Glaister, S., Grayling, T., Hallam, K., Graham, D. and Anderson, R., (2002) Streets Ahead: safe and liveable streets for children, IPPR, ISBN:1- 86030-207-6

Goldstein, H (1990) Problem based policing, Sage Publishing, London

Goodchild, M. (1986) Spatial Auto-correlation, Catmog Series 47

Graham, D. and Glaister, S. and Anderson, R. (2002) Child pedestrian casualties in England: the effect of area deprivation, Centre for Transport Studies, Imperial College London

Graham, D. and Glaister, S. and Anderson, R. (2003) The effects of area deprivation on the incidence of child and adult pedestrian casualties in England, Accident Analysis and Prevention, Vol. 37 No. 1 pp. 185-191

Greenwood, M. and Woods, H. (1919) On the incidence of industrial accidents upon individuals, with special reference to multiple accidents, Report on the Industrial Fatigue Research Board, Her Majesty's Stationary Office, No. 4 pp. 1-28

Greenwood, M. and Yule, G. (1920) An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attack of disease or repeated accidents, Journal of the Royal Statistical Society, No. 83 pp. 255-279

Greenwood, M. (1951) Accident proneness, Biometrika, Vol. 37 No. 1-2 pp. 24-29

Greibe, P. (2003) Accident prediction models for urban roads, Accident Analysis and Prevention, Vol. 35 No. 2 pp. 273-285

Griffith, D. A. (1987) Spatial Autocorrelation: A Primer, Washington, DC: Association of American Geographers, Resource Publications in Geography

Gupta, S.S. and J.C. Hsu (1980) Subset Selection Procedures With Application to Motor Vehicle Fatality Data in a Two-Way Latout Technometrics, Vol. 22 No. 4 pp. 543-546.

Guptill, S. (1989) Inclusion of accuracy data in feature based, object orientated data model, Accuracy in spatial databases, Taylor and Francis, London

Hakamies-Blomqvist, L. and Siren, A. (2003) Deconstructing a gender difference: Driving cessation and personal driving history of older women, Journal of Safety Research, Vol. 34 No. 4 pp. 383-388

Haepers, A. S. and Pocock, P. T. (1993) Road traffic accident involving children from ethnic minorities Report for Leeds County Council

Haggett, P. A. & Cliff, D. & Frey, A.E. (1977) Locational Analysis In Human Geography, Second Edition, Arnold, London

Hamerslag, R. and Roos, J. and Kwakernaak, M. (1982) Analysis of accidents in traffic situations by means of multiproportional weighted poisson model, Transportation Research Board, No. 847, Washington, USA

Harris, R. (2003) The diversity of diversity: is there still a place for small area classifications?, Area, Vol. 33 No. 3 pp. 329-336

Hassan, Y. and Barker, D. (1999) The impact of unseasonable or extreme weather on traffic activity within Lothian region, Scotland, Journal of Transport Geography, Vol. 7 No. 2 pp. 209-213

Hasselberg, M. and Laflamme, L. (2005) The social patterning of injury repetitions among young car drivers in Sweden, Accident Analysis and Prevention, Vol. 37 No. 1 pp. 163-168

Hauer, E. and Ahlin, , F. and Bowser, J. (1982) Speed enforcement and speed choice, Accident Analysis and Prevention, Vol. 14 No. 4 pp. 267-278

Hauer, E. and B.N. Persaud (1987) How to estimate the safety of rail-highway grade crossing and the effects of warning devices Transportation Research Record Vol.1114 pp. 131-140.

Hauer, E. (1997) Observational before-after studies in road safety, Pergamon, Oxford

Henning-Hager, U. (1986) Urban development and road safety, Accident Analysis and Prevention, Vol. 18 No.2 pp.135-145

Haynes *et al* (2005) Geographical distribution of road traffic deaths in England and Wales: place of accident compared with place of residence, Journal of Public Health, Vol. 27 No. 1 pp. 107-111

Herslund, M. and Jorgensen, N. (2003) Looked-but-fail-to-see-errors in traffic, Accident Analysis and Prevention, Vol. 35 No. 6 pp. 885-891

Hewson, P. (2005) Epidemiology of child pedestrian casualty rates: Can we assume spatial independence?, Accident Analysis and Prevention, Vol. 37 No. 4 pp. 651-659

Hing, J. and Stamatiadis, N. and Aultman-Hall, L. (2003) Evaluating the impact of passengers on the safety of older drivers, Journal of Safety Research, Vol. 34 No. 4 pp. 343-351

HMIC Thematic Inspection Report (1998) Road policing Traffic Report

Home Office Research Study 124 (1991) Traffic Policing in Changing Times, HMSO

Hummel, T. (2001) Land use planning in Safer transportation network planning, SWOV Institute for Road Safety Research, The Netherlands

Hunsaker, C.T. and Goodchild, M.F. and Friedl, M.A. and Case, E.J. (2001) (eds) Spatial Uncertainty in Ecology: Implications for Remote Sensing and GIS Applications, Springer, New York

Ibrahim K. and Silcock D. T. (1992) The accuracy of accident data, Traffic Engineering and Control, Vol. 33 pp.492-497

Ivan *et al* (1999) Differences in causality factors for single and multi-vehicle crashes on two lane roads, Accident Analysis and Prevention, Vol. 31 No. 6 pp. 695-704

Ivan, J. and Wang. C. and Bernado, N. (2000) Explaining two-lane highway crash rates using land use and hourly exposure, Accident Analysis and Prevention, Vol. 32 No. 6 pp. 787-795

James, H. (1991) Under-reporting of road traffic accidents, Traffic Engineering and Control, Vol. 27 pp. 574-583

Janssen, W. and Tenkink, W. (1987) Risk homeostasis theory and its critics: time for agreement, TNO Intitute of Perception, The Netherlands

Joly, M-F., Boubeau, R., Bergeron, J., and Messier, S. (1992) Analytical Approach to the Identification of Hazardous Road Locations: A Review of the Literature. Centre de recherche sur les transports, Universite´ de Montre´al, CRT publication No. 815.

Jones, A. and Jorgensen, S. (2003) The use of multilevel models for the prediction of road accident outcomes, Accident Analysis and Prevention, Vol. 35 No. 1 pp. 59-69

Jovanis, P. and J. Delleur (1983) Exposure-based analysis of motor vehicle accidents Transportation Research Record No. 910 pp. 1-7.

Judge K, Benzeval M. (1993) Health inequalities: new concerns about the children of single mothers. BMJ Vol. 306 pp. 677-80

Julien, A. and Carré, J. (2002) Risk exposure during pedestrian journeys, Recherche Transports Sécurité, Vol. 76 pp. 173-189

Keall, M. (1995) Pedestrian Exposure to risk of road accident in New Zealand, Accident Analysis and Prevention, Vol. 27 No. 5 pp. 729-740

Kennedy, M. (2002) The Global Positioning System and GIS: An Introduction, CRC Press Inc, USA

Keskinen, E. and Ota, H. and Katila, A. (1998) Older drivers fail in intersections: speed discrepancies between older and younger male drivers, Accident Analysis and Prevention, Vol. 30 No. 3 pp. 323-330

Klein, D. (1980) Societal influences on childhood accidents, Accident Analysis and Prevention, Vol. 12 pp. 275-281

Kmet, L. and Brasher, P. and Macarthur, C. (2003) A small area study of motor vehicle crash fatalities in Alberta, Canada, Accident Analysis and Prevention, Vol. 35 No. 2 pp. 177-182

Knox, P. and Pinch, S. (2000) Urban Social Geography: An introduction, 4th Edition, Prentice Hall, New York

Kposowa, A. and Adams, M. (1998) Motor vehicle crash fatalities: The effects of race and marital status, Applied Behavioural Science Review, Vol. 6 No. 1 pp 69-91

Kulmala, R. (1995) Safety at rural three- and four-arm junctions Technical Research Centre of Finland (VTT), Espoo, Finland

Laflamme *et al* (2005) Car safety and social differences in traffic injuries among young adult drivers: a study of two-car injury-generating crashes in Sweden, Safety Science, Vol. 43 No. 1 pp 1-10

Lalloo, R. and Sheiham, A. and Nazroo, J. (2003) Behavioural characteristics and accidents: findings from the Health Survey for England, 1997, Accident Analysis and Prevention, Vol. 35 No. 5 pp. 661-667

Laapootti, S. and Keskinen, E. (2004) Has the difference in accident patterns between male and female drivers changed between 1984 and 2000?, Accident Analysis and Prevention, Vol. 36 No. 4 pp. 577-584

LaScala, E. and Gerber, D. and Gruenewald, P. (2000) Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis, Accident Analysis and Prevention, Vol. 32 No. 5 pp. 651-658

LaScala, E. and Gruenewald, P. and Johnson, F. (2004) An ecological study of the locations of schools and child pedestrian injury collisions, Accident Analysis and Prevention, Vol. 36 No. 4 pp. 569-676

Lassarre, S. and Thomas, I. (2005) Exploring road mortality ratios in Europe: national versus regional realities, Journal of the Royal Statistical Society A, Vol. 168 No. 1 pp. 127-144

Latimer (1992) Effects of increased auto safety belt use levels on fatalities Risk Analysis, pp. 449-454.

Lawson I. (1990) Determining socio economics as a influence on road accidents: A case study of Ohio, USA, Safety Science Vol. 12 No.3

Lawson, I. & Edwards, P.J. (1991) The involvement of ethnic minorities in road accidents: Data from three studies of young pedestrian casualties, Traffic Engineering and Control, Vol. 32 pp. 12-19

Lee, S. (1986) Accidents to Asians Research Note from Road Safety Unit of Rochdale Metropolitan Borough Council, (unpublished).

Lenguerrand, E. and Martin, J. and Laumon, M (2006) Modelling the hierarchical structure of road crash data—Application to severity analysis, Accident Analysis and Prevention, Vol. 38 No. 1 pp. 43-53

Leveson, N. (2004) A new accident model for engineering safer systems, Safety Science, Vol. 42 No. 4 pp. 237-270

Levine, N. and Kim, K. and Nitz, L. (1995) Daily fluctuations in Honolulu motor vehicle accidents, Accident Analysis and Prevention, Vol. 27 No. 6 pp. 785-796

Lloyd, D. (2005) Uncertainty in town centre definition, PhD Thesis, Centre for Advanced Spatial Analysis, London

Lonero, L. (1998) Risk mentality: Why drivers take the risks they do (and how we can help), 8th World Traffic Safety Symposium, New York, USA

Longley, P.A., Webber, R. (2003) Geodemographic Analysis of Similarity and Proximity: Their Roles in the Understanding of the Geography of Need. In P.A. Longley and M. Batty (eds) Advanced Spatial Analysis: The CASA Book of GIS, pp 233-266. Redlands: ESRI Press.

Longley, P. (2005) Geographical Information Systems: a renaissance of geodemographics for public service delivery, Progress in Human Geography, Vol. 29 No. 1 pp. 57-63

Longley *et al* (2005) Geographic Information Systems and Science, John Wiley and Sons, Chichester

Lu, T. and Chou, M. and Lee, M. (2000) Regional mortality from motor vehicle traffic injury: relationships among place-of-occurrence, place-of-death, and place-of-residence, Accident Analysis and Prevention, Vol. 32 No. 1 pp. 65-69

Lubans, V.A. and Edgar, J.M. (1979) Policing by Objectives, Hartford, CT.: Social Research Foundation

MacIntyre, S. and MacIver, S. and Sooman, A. (1993) Area, Class and Health: Should we be Focusing on Places or People? Journal of Social Policy, Vol. 22 pp. 213-234.

Maher, M. and Summersgill, I. (1996) A comprehensive methodology for the fitting of predictive accident models, Accident Analysis and Prevention, Vol. 28 No. 3 pp 281-296

Martin, D. (1996) An assessment of surface and zonal models of population International Journal of Geographical Information Systems Vol.10 pp. 973-989

Martin, D. and Tate, N. J. and Langford, M. (2000) Refining population surface models: experiments with Northern Ireland Census Data Transactions in GIS Vol. 4 pp. 343-360

Martin, J. (2002) Relationship between crash rate and hourly traffic flow on interurban motorways, Accident Analysis and Prevention, Vol. 34 No. 5 pp. 619-629

Massie, D. and Campbell, K. and Williams, A. (1995) Traffic Accident Involvement rates by driver age and gender, Accident Analysis and Prevention, Vol. 27 No. 1 pp. 73-87

Maycock, G. and R.D. Hall (1984) R.D., Accidents at 4-arm roundabouts Reports 1120 Crowthorne, U.K., Transport and Road Research Laboratory

Mayou, R. and Byrant, B. (2003) Consequences of road traffic accidents for different types of road user, International Journal of the Care of the Injured, Vol. 34 pp. 197-202

McGuigan, D.R.D. (1981) The use of relationships between road accidents and traffic flow in black-spot identification Traffic Engineering and Control Vol. 22 No. 8-9 pp. 448-453

McKenna, F. and Waylen, A. and Burkes, M. (1998) Male and female drivers: how different are they?, AA Foundation for Road Safety Research

Mekky, A. (1985) Effects of rapid increase in motorization levels on road fatality rates in some rich developing countries, Accident Analysis and Prevention, Vol. 17 No. 2 pp. 101-109

Miaou, S.P. (1994) The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions Accident Analysis and Prevention Vol. 26 No.4 pp. 471- 482

Millot, M. (2004) Urban growth, travel practises and evolution of road safety, Journal of Transport Geography, Vol. 12 No. 3 pp. 207-218

Moller, M. (2004) An explorative study of the relationship between lifestyle and driving behaviour among young drivers, Accident Analysis and Prevention, Vol. 36 No. 6 pp. 1081-1088

Mountain, L. and Fawaz, B. and Jarrett, D. (1996) Accident prediction models for roads with minor junctions, Accident Analysis and Prevention, Vol. 28 No. 6 pp. 695-707

Murray, A. (1998) The home and school background of young drivers involved in traffic accidents, Accident Analysis and Prevention, Vol. 30 No. 2 pp. 169-182

Nader, R. (1965) Unsafe at any speed: The designed-in dangers of the American automobile, Grossman Publishers, New York

Nassar, S. and Saccomanno, F. and J. Shortreed (1994) Road accident analysis: A micro-level approach Canadian Journal for Civil Engineers Vol. 21 No. 5 pp. 847-855.

Natanen, R. & Summala, H. (1974) A model for the role of motivational factors in drivers' decision-making, Accident Analysis and Prevention, No. 6 pp.243-26

Natanen, R. & Summala, H. (1976) Road-user behaviour and traffic accidents Amsterdam and New York: North-Holland/American Elsevier

Naveh, E. & Marcus, A. (2002) A Re-Assessment of Road Accident Data-Analysis Policy, Center for Transportation Studies, University of Minnesota

Neilson, I. and Condon, R. (2000) Desirable improvements in road accidents and related data, PACTS Research Paper, No. 1

Newold, E. (1926) A contribution to the study of the human factors of causation of accidents, Industrial Fatigue Research Board, Her Majesty's Stationary Office, No 34

Newold, E. (1927) Practical applications of the statistics of repeated events particularly industrial accidents, Journal of the Royal Statistical Society, Vol. 90 pp. 487-547

Ng, K. and Hung, W. and Wong, W. (2002) An algorithm for assessing the risk of traffic accident, Journal of Safety Research, Vol. 33 No. 3 pp. 387-410

Nguyen, T. N. (1991) Identification of Accident Blackspot Locations, an Overview. VIC Roads/Safety Division, Research and Development Department, Discussion Paper (DP/91/4), Victoria, Australia

Nicholson, A. (1999) Analysis of spatial distributions of accidents, Safety Science, Vol. 31 No. 1 pp. 71-91

Noland, R. (2003) Traffic fatalities and injuries: the effect of changes in infrastructure and other trends, Accident Analysis and Prevention, Vol. 35 No. 4 pp. 599-611

Noland, R. and Quddus, M. (2000) An analysis of pedestrian and bicycle casualties using panel data, Centre for Transport Studies, Imperial College of Science, London

Noland, R. and Oh, L. (2004) The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data, Accident Analysis and Prevention, Vol. 36 No. 4 pp. 525-532

Noland, R. and Quddus, M. (2004) A spatially disaggregate analysis of road casualties in England, Accident Analysis and Prevention, Vol. 36 No. 6 pp. 973-984

Odland, J. (1988) Spatial Autocorrelation, Sage Publications, USA

Ogilvie-Smith, A. and Downey, A and Ransom, E. (1994) Traffic Policing: Activity and Organisation, Police Research Series, Paper 12

Openshaw, S. and Taylor, P. (1979) A million or so correlation coefficients: three experiments on the modifiable areal unit problem In: Statistical Applications in the Spatial Sciences. pp. 127-144. Edited by N. Wrigley. Pion, London.

Openshaw, S. (1984) The modifiable unit problem Norwich, GeoBooks Concepts and Techniques in Modern Geography No. 38

Openshaw, S. (1995) Developing automated and smart spatial pattern exploration tools for geographical information systems applications The Statistician, 44, 3–16.

Openshaw, S. & Alvanides, S. (1999) Applying geocomputation to the analysis of spatial distributions. In: P.A. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind, eds. *Geographical Information Systems: Principles, Techniques, Management and Applications*. 2nd edition. Chichester: John Wiley, 267-82

Openshaw, S., and Turton, I. (2001) Geographical analysis machine on the Internet, Available on the World Wide Web http://www/ccg.leeds.ac.uk/smart/gam/gam.html (Accessed 29th July 2003)

Oppe, S. (1979) The use of multiplicative models for analysis of road safety data Accident Analysis and Prevention Vol.11 pp. 101-115

Oppe, S (1993) Possibilities and limitations of accident analysis, 6th ICTCT workshop Salzburg, Germany

Ossenbruggen, P. and Pendharkar, J. and Ivan, J. (2001) Roadway safety in rural and small urbanized areas, Accident Analysis and Prevention, Vol. 33 No. 4 pp. 485-498

Oxley, J., Charlton, J., Fildes, B., Koppel, S., Scully, J., Congiu, M. & Moore, K. (2005) Crash risk of older female drivers, Report No. 245, Monash University Accident Research Centre

Pasquier, M. & Quek, C. & Tung, W.L. & Chen, D. & Yep, T.M. (2002) Fuzzylot II: A novel soft computing approach to the realisation of autonomous driving manoeuvres for intelligent vehicles, Seventh International Conference on Control, Automation, Robotics and Vision, Singapore

Pau, M. and Angius, S. (2001) Do speed bumps really decrease traffic speed? An Italian experience, Accident Analysis and Prevention, Vol. 33 No. 5 pp. 585-597

Petch, R. and Henson, R. (2000) Child road safety in the urban environment, Journal of Transport Geography, Vol. 8 No. 3 pp. 197-211

Perchonok, K. (1972) Accident cause analysis. CAL Report. ZM-5010-V-3, Cornell Aeronautical Laboratory, Inc.

Persaud, B. (1990) Black spot identification and treatment evaluation The Research and Development Brach, Ontario, Ministry of Transportation.

Pietro P 2001 High Risk road users, NRMA ACT Road Safety Trust, Australia

Pulugurtha, S. and Vanjeeswaren, K. and Nambisan, S. (2003) Development of Criteria to Identify Pedestrian High Crash Locations in Nevada, Quarterly Progress Report, University of Nevada, USA

Ratcliffe, J. (2000) Aoristic analysis: the spatial interpretation of unspecific temporal events International Journal of Geographical Information Science, Vol. 14 No. 7 pp. 669-679.

Retting *et al* (1995) Classifying urban crashes for countermeasure and development, Accident Analysis and Prevention, Vol. 27 No. 3 pp. 283-294

Roberts, I. and Power, C. (1996) Does the decline in child injury mortality vary by social class? A comparison of class specific mortality in 1981 and 1991. British Medical Journal, Vol. 313 pp. 784-786

Robinson, P., and Nolan, T. (1997) Paediatric Slow-Speed Non-Traffic Fatalities: Victoria, Australia, 1985-1995. Accident Analysis and Prevention, Vol. 29 pp.731-737.

Ross, A., and Bagunley, C. and Hills, B. and` McDonald, M. and Silcock, D. (1991). Towards Safer Roads in Developing Countries - a Guide for Planners and Engineers, TRL, Crowthorne, UK.

Rothe, J.P (2002) Driving Lessons: Exploring system that make traffic safer, University of Alberta Press, Canada

Rumar, K. (1999), The 1st European Transport Safety Lecture: Transport Safety Visions, Targets and Strategies: Beyond 2000, European Transport Safety Council, Brussels

Ryan, G. and Legge, M. and Rosman, D. (1998) Age related changes in drivers' crash risk and crash type, Accident Analysis and Prevention, Vol. 30 No. 3 pp. 379-387

Sabel C E, Gatrell A C, Löytönen M, Maasilta P and Jokelainen M.(2005) Modelling exposure opportunities: estimating relative risk for motor neurone disease in Finland, Social Science and Medicine, Vol. 50 pp. 1121 - 1137

Sabel, C. (2006) kernel density estimation as a spatial-temporal data mining tool: exploring road traffic accident trends, GISRUK 2006, University of Nottingham

Sabey, B. E. and Staughton, G. C. (1975) Interacting roles of road environment, vehicle and road user in accidents. Paper presented at the 5th International Conference on the International Association for Accident and Traffic Medicine, London

Saccomanno, F. and Buyco, C. (1988) Generalized loglinear models of truck accident rates Transportation Research Record Vol. 1172 pp. 23-31.

Saccomanno, F. and Fu, L. (2000) Gis-Based Integrated model foR road accident analysis and prediction, Department of Civil Engineering, University of Waterloo

Sawalha, Z. & Sayed, T. (2001) Evaluating Safety of Urban Arterial Roadways, Journal of Transportation Engineering, Vol. 127 No. 2 pp.151-158

Sawalha, Z. and Sayed, T. (2006) Transferability of accident prediction models, Safety Science, Vol. 44 No. 3 pp. 209-219

Sayed, T. and Abdelwahab, W. and Navin, F. (1995) Identifying accident-prone locations using fuzzy pattern recognition, Journal of Transportation Engineering, Vol. 121 No. 4 pp. 352-358

Scheiner, J. and Kasper, B. (2003) Lifestyles, choice of housing location and daily mobility: the lifestyle approach in the context of spatial mobility and planning, International Social Science Journal, Vol. 55 No. 3 pp.319-332

Schneider, R. and Khattak, A. and Rzynar, R. (2001) Identifying factors associated with pedestrian crash risk through spatial analysis and perception-based modelling, Transportation Research Board

Shankar, V., Mannering, F., and W. Barfield (1995) Effect of roadway and environmental factors on rural freeway accident frequencies Accident Analysis and Prevention Vol. 27 No.3 pp. 371-389

Silcock, D. T. and Smyth, A. W. (1985) Methods of Identifying Accidents Blackspots Transport Operations Research Group, Department of Civil Engineering, University of Newcastle upon Tyne. TORG Research Report

Simons-Morton, B. and Lerner, N. and Singer, J. (2005) The observed effects of teenage passengers on the risky driving behaviour of teenage drivers, Accident Analysis and Prevention, Vol. 37 No. 6 pp. 973-982

Silverman, B. (1986) Density Estimation for Statistics and Data Analysis, 1st edition, Chapman and Hall, London

Sjogren *et al* (1996) Differences between older and younger drivers; characteristics of fatal car crashes and driver injuries, Safety Sceince, Vol. 23 No. 1 pp. 63-77

Sjogren *et al* (1996) Role of disease in intiating the crashes of fatally injured drivers, Accident Analysis and Prevention, Vol. 28 No. 3 pp. 307-314

Sleet, D. and Branche, C. (2004) Road safety is no accident, Journal of safety research, Vol. 35 No. 2 pp. 173-174

Smeed, R.J. (1968) Variations in the pattern of accident rates in different countries and their causes, Traffic Engineering and Control Vol.10 pp. 364–71.

Smeed, R.J. and Jeffcoate, G.O. (1970) Effects of changes in motorisation in various countries on the number of road fatalities, Traffic Engineering and Control, Vol.12 pp. 150–51.

Social Exclusion Unit (2003) Making the Connections: Final Report on Transport and Social Exclusion, Office for the Deputy Prime Minister

Standish, W. (2003) Categorising risk in road safety Journal of Safety Science Vol. 23 pp.987-999

Steenberghen *et al* (2004) Intra-urban location and clustering of road accidents using GIS: a Belgian example, International Journal of Geographical Information Science, Vol. 18 No. 2 pp. 169-181

Sullivan, J. and Flannigan, M. (2002) The role of ambient light level in fatal crashes: inferences from daylight saving time transitions, Accident Analysis and Prevention, Vol. 34 No. 4 pp. 487-498

Summala, H. (1985). Modelling driver behaviour: A pessimistic prediction? In Evans, L. & Schwing, R. C. (Eds), Human behaviour and traffic safety (pp. 43-65). New York: Plenum.

Summala, H. (1988). Risk control is not risk adjustment: The zero-risk theory of driver behavior and its implications, Special Issue: Risky decision-making in transport operations. Ergonomics Vol. 31 pp. 491-506.

Summala, H. (1996) Accident risk and driver behaviour, Safety Science, Vol. 22 No. 1-3 pp. 103-117

Tarko, A. and Tracz, M. (1995) Accident prediction models for signalized crosswalks, Safety Science, Vol. 19 No. 2-3 pp. 108-119

Taylor, J. and Deane, F. and Podd, J. (2000) Determining the focus of driving fears, Journal of Anxiety Disorders, Vol. 14 No. 5 pp. 453-470

Thomas, I. (1996) Spatial Data Aggregation: Exploratory analysis of road accidents, Accident Analysis and Prevention, Vol. 28 No. 2 pp. 251-264

Thurstain-Goodwin, M. and Unwin, D. (2000) Defining and delineating the central areas of towns for statistical monitoring using continuous surface representations, Transactions in GIS, Vol. 4 No. 4 pp. 305-317

Tight *et al* (1998) Casualty reduction or danger reduction: conflicting approaches or means to achieve the same ends?, Transport Policy, Vol. 5 No. 3 pp.185-192

Tobler, W.(1970) A Computer Movie Simulating Urban Growth in the Detroit Region, Economic Geography, Vol. 46 No. 2 pp. 234-240

Torrens, P. M. and Alberti M.(2000) Measuring Sprawl, CASA (Centre for Advanced Spatial Analysis) Working Papers 27 (available at: http://www.casa.ucl.ac.uk)

Transport for London (2000) Roads Policing strategy

Transport for London (2004) Accident and casualties on London's roads 2003, London Accident Analysis Unit, London

Transport for London (2005) Bus and coach casualties in Greater London London Accident Analysis Unit, London

Transport for London (2001) Tackling the road safety implications of disadvantage, London Accident Analysis Unit, London

Transport for London (2004) Goods vehicle accidents and casualties in Greater London, London Accident Analysis Unit, London

Transport for London (2003) Older car occupant casualties in Greater London, London Accident Analysis Unit, London

Transport for London (2003) Older pedestrian casualties in Greater London, London Accident Analysis Unit, London

Transport for London (2005) Pedal cyclist casualties in Greater London, London Accident Analysis Unit, London

Transport for London (2005) Pedestrian casualties in Greater London, London Accident Analysis Unit, London

Transport for London (2004) Levels of accident risk in Greater London, London Road Safety Unit, London

Transport for London (2001) London's road safety plan, London Road Safety Unit, London

Transport for London (2004) Powered two wheeled user casualties in Greater London, London Accident Analysis Unit, London

Transport for London (2005) Collisions involving young car drivers in Greater London, London Accident Analysis Unit, London

Treat, J. R., Tumbas, N. S., McDonald, S .T., Shinar, D., Hume, R. D., Mayer, R. E., Stanisfer, R. L., and Castillan, N. J. (1977) Tri-level study of the causes of traffic accidents. Report No. DOT-HS-034-3-535-77, Indiana University

Tunaru, R. (1999) Hierarchical Bayesian models for road accident data Traffic Engineering and Control, Vol. 40 No.6 pp. 318-324

Ulleberg, P. (2002) Personality subtypes of young drivers. Relationship to risk-taking preferences, accident involvement, and response to a traffic safety campaign, Transportation Research Part F, Vol. 4 pp. 279-297

Unwin, A. R. (1996) Exploratory spatial analysis and local statistics, Computational Statistics, Vol. 11 pp. 387-400

Valent, F. (2002) Risk factors for fatal road traffic accidents in Udine, Italy, Accident Analysis and Prevention, Vol. 34 No. 1 pp. 71-84

Van Eck, N.J &, Frasincar, F. & van den Berg, J. (2006) Visualizing Concept Associations Using Concept Density Maps, Tenth International Conference on Information Visualisation (IV'06), pp. 270-275

Van den Bossche F. and Wets G. and E. Lesaffre (2003) A Bayesian hierarchical approach to model the rank of hazardous intersections for bicyclists using the Gibbs sampler In Proceedings Transportation Research Board (CD-ROM), Washington, USA, 11-16 January

Venables, W. and Ripley, B. (1997) Modern Applied Statistics with S-Plus, Springer

Vistisen, D. (2002) Models and methods for hot spot safety work, PhD thesis, Denmark.

Voas, D. and Williamson, P. (2001) The diversity of diversity: a critique of geodemographic classification, Area, Vol. 33 No. 1 pp. 63-76

Webber, R. (2003) Designing geodemographic classifications to meet contemporary business needs, Interactive Marketing, Vol. 5 No. 3 pp. 219-237

Webber, R. (2004) The relative power of geodemographics vis a vis person and household level demographic variables as discriminators of consumer behaviour, CASA Working Paper Series, No. 84

Webber, R. (2004) Demographic and deprivation rations: examples of their use in understanding underlying spatial patterns in social phenomena, CASA Working Paper, No. 86

White, D. and Raeside, R. and Barker, D. (2000) Road accidents and children living in disadvantaged areas: A literature review, Scottish Executive Central Research Unit

Whitelegg, J. (1987) A geography of road traffic accidents, Transactions of the Institute of British Geographers, Vol. 12 pp. 161-176

Wilde, G.J.S. (1976) Social Interaction Patterns in Driver Behaviour; an Introductory Review. Human Factors, No. 18, pp. 477-492

Wilde, G.J.S. (1982). The theory of risk homeostasis: Implications for safety and health, Risk Analysis, Vol. 2, pp. 209–225.

Wilson, E. M. and Lipinski, M. E. (2004) Road Safety Audits: A Synthesis of Highway Practice, National Cooperative Highway Research Program (NCHRP) Synthesis 336: Road Safety Audits (http://trb.org/publications/nchrp/nchrp_syn_336.pdf)

World Health Organisation (2004) World Report on road traffic injury prevention, World Health Day Publication

Williams, A. (2003) Teenage drivers: patterns of risk, Journal of Safety Research, Vol. 34 No. 1 pp. 5-15

Williams, A. and Shabanova, V. (2003) Responsibility of drivers, by age and gender, for motor-vehicle crash deaths, Journal of Safety Research, Vol. 34 No. 5 pp. 527-531

Woods, D.P. and C.K. Simms (2002) Car size and injury risk: a model for injury risk in frontal collisions Accident Analysis and Prevention Vol. 34 pp. 93-99.

Yule, G. and Kendall, M. (1950) An Introduction to the Theory of Statistics, Griffin and Co, London

Zegeer, C.V. *et al* (1990) Cost effective geometric improvements for safety upgrading of horizontal curves. Chapel Hill, University of North Carolina.

Zhang, J. and Goodchild, M. (2002) Uncertainty in Geographical Information, Taylor and Francis, New York